Joint Source, Channel and Space-time Coding of Progressive Sources in MIMO Systems

Meesue Shin, *Student Member, IEEE*, Laura Toni, *Member, IEEE*, Sang-Hyo Kim, *Member, IEEE*, and Seok-Ho Chang, *Member, IEEE*

Abstract

The optimization of joint source and channel coding for a sequence of numerous progressive packets is a challenging problem. Further, the problem becomes more complicated if the space-time coding is also involved with the optimization in a multiple-input multiple-output (MIMO) system. This is because the number of ways of jointly assigning channels codes and space-time codes to progressive packets is much larger than that of solely assigning channel codes to the packets. We are unaware of any feasible and complete solution for such optimization of joint source, channel, and space-time coding of progressive packets. This paper applies a parametric approach to address that complex joint optimization problem in a MIMO system. We use the parametric methodology to derive some useful theoretical results, and then exploit those results to propose an optimization method where the joint assignment of channel codes and space-time codes to the packets can be optimized in a packet-by-packet manner. As a result, the computational complexity of the optimization is exponentially reduced, compared to the conventional exhaustive search. The numerical results show that the proposed method significantly improves the peak-signal-to-noise ratio performance of the rate-based optimal solution in a MIMO system.

Index Terms

Diagonal Bell Labs space-time architecture (D-BLAST), joint source and channel coding, multiple-input multiple-output (MIMO), optimization, orthogonal space-time block codes (OSTBC), outage probability, progressive source, space-time codes, vertical BLAST (V-BLAST)

M. Shin and S.-H. Chang are with the Department of Computer Science and Engineering, Dankook University, Yongin 448-701, Korea (e-mail: 32101705@dankook.ac.kr; seokho@dankook.ac.kr). L. Toni is with the Department of Electronic and Electrical Engineering at University College London (UCL), Gower Street, London, WC1E 7JE (e-mail: l.toni@ucl.ac.uk). S.-H. Kim is with the College of Information and Communication Engineering, Sungkyunkwan University, Suwon 16419, Korea (e-mail: iamshkim@skku.edu).

I. Introduction

In recent years, there has been significant demand for the transmission of multimedia services over wireless channels, and this has motivated intense research into cross-layer optimization design [1], [2], which is particularly important for mobile radio channels that exhibit time-variant channel-quality fluctuations. Progressive sources, such as embedded images or scalable video [3], [4], employ a manner of transmission such that the quality of the decoded source improves when the number of successfully received bits increases. However, such advances in source coders have made the source bitstream very susceptible to impairments due to mobile fading channels.

Multiple-input multiple-output (MIMO) technology is an important advance in wireless communications in terms of the link reliability and data rate. Spatial diversity schemes, such as orthogonal space-time block codes (OSTBC) [5], improve reliability by extracting the diversity gain to combat signal fading from the channels. The OSTBC is an important class of linear space-time codes, in that it acquires the full diversity of channels with a very simple linear receiver. Spatial multiplexing schemes use a layered approach to increase the data rate [6]. One popular example is the vertical Bell Laboratories layered space-time (V-BLAST) architecture, where independent data signals are transmitted over antennas to increase the data rate, but full spatial diversity is usually not achieved.

In this paper, we study the optimization of joint source, channel, and space-time coding of progressive sources in such a MIMO system. Progressive source encoders produce data with gradual differences in the importance of their bitstreams. We consider the system where the bitstream is taken from the progressive source encoder, and is transformed into a sequence of L packets. Such a system is depicted in Fig. 1. Each of those L progressive packets can be encoded with different channel codes and modulations in a similar way to the works in [7]–[14]. Further, each packet can be encoded with different space-time codes [15]–[20], to achieve the best end-to-end performance as measured by the expected distortion of the source. We assume that all the encoded packets have the same time duration, T_{pkt}, and the same signal bandwidth, W_{pkt}. We let u_i denote the spectral efficiency (bits/s/Hz) of the ith packet that has been encoded by a given channel code and modulation ($1 \leq i \leq L$); u_i is determined by the code rate of the channel

code, and by the alphabet size of the modulation. We let v_i denote the spatial multiplexing rate[1] of the ith packet that has been encoded by a space-time code. Then, the number of information (or source) bits in the ith packet is expressed as $u_i v_i T_{\rm pkt} W_{\rm pkt}$. As we increase either the spectral efficiency, u_i, or the spatial multiplexing rate, v_i, the variance of the quantization error from the source coder decreases, but the probability of the packet error, caused by signal fading and noise from the channels, increases.

Let $N_{\rm se}$ denote the number of candidate spectral efficiencies considered in a system. The number of possible assignments of $N_{\rm se}$ spectral efficiencies to a sequence of L progressive packets is $N_{\rm se}^L$, which exponentially grows as L increases. As an example, for the transmission of a 512×512 progressive image with a rate of 1 bit-per-pixel (bpp), a sequence of $L = 128$ packets is considered in [7]. Further, in the MIMO system depicted in Fig. 1, if each packet can be encoded with different space-time codes (e.g., V-BLAST, OSTBC, and two-layer diagonal BLAST (D-BLAST)), which offer different spatial multiplexing rates, the assignment of spatial multiplexing rates as well as spectral efficiencies to L packets yields a more complicated optimization problem. To address this matter, for a single-input single-output (SISO) system, there have been many studies about the optimal assignment of spectral efficiencies (or, equivalently, the channel code rates, with the alphabet size of the modulation being fixed) to a sequence of progressive packets [7]–[14]. For a MIMO system, however, those studies do not immediately indicate how to jointly assign spectral efficiencies and spatial multiplexing rates to progressive packets. Although the previous works in [7]–[14] on joint progressive source and channel coding were presented several years ago, to our knowledge, none of those works have been successfully extended to the case where space-time coding is also involved with the optimization.

To address that complex optimization problem in a MIMO system, there have been some researches regarding the optimal assignment of space-time codes to a sequence of progressive packets [18]–[20]. Those works have focused on significantly reducing the number of ways to assign space-time codes to progressive packets, but do not provide a complete solution for how to jointly assign the spectral efficiencies and spatial multiplexing rates to the packets. In [18]–[20], it was shown that performance of the progressive transmission in MIMO systems

[1]Spatial multiplexing rate is defined as the ratio of the number of symbols packed within a space-time codeword to its time duration expressed in terms of the number of symbols. As an example, for two transmit antennas, the Alamouti code (i.e., OSTBC) and V-BLAST achieve spatial multiplexing rates of 1 and 2, respectively.

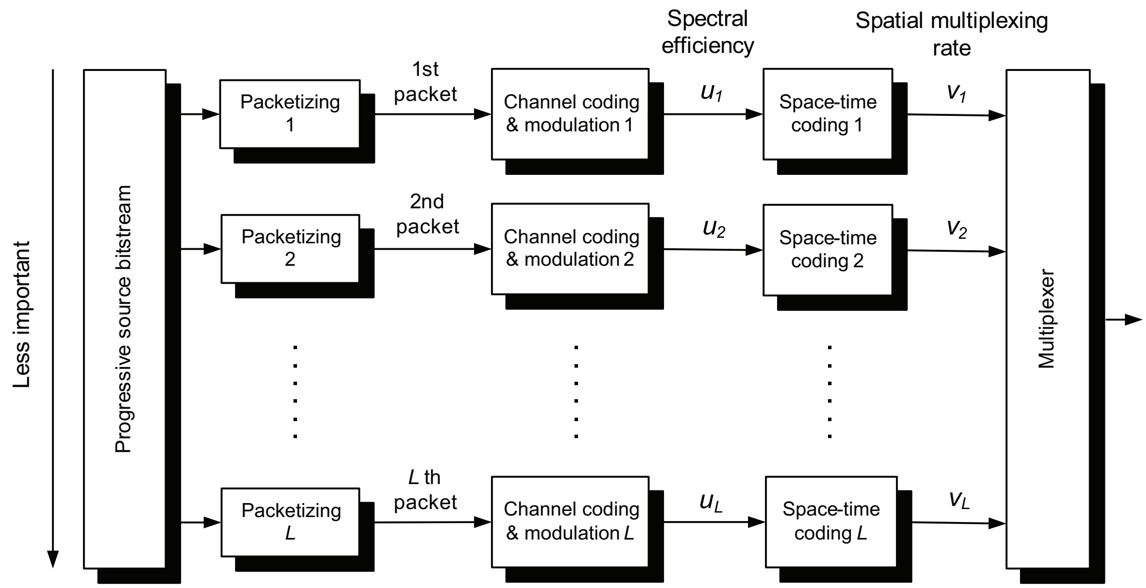


Fig. 1. A progressive source transmission in a MIMO system. u_i denotes the spectral efficiency (bits/s/Hz), and v_i denotes the spatial multiplexing rate of the ith packet ($1 \leq i \leq L$).

is sensitive to the way space-time codes are assigned to a sequence of packets. However, to our knowledge, solutions for such joint optimization problem in a MIMO system have not been presented yet in the literature. For example, in [16], [17], [21]–[24], the transmission of progressive images or scalable video over MIMO systems have been studied; specifically, the work in [16] showed the performance improvement when applying different space-time codes to two unequally important source layers (base and enhancement layers). However, those works did not study joint optimization for a series of progressive packets, in such a manner that the works in [7]–[14] studied for a SISO system.

In this paper, we use a parametric approach to address that complex joint optimization problem in a MIMO system. Specifically, we employ a parametric model of the operational distortion-rate function of the source, which is based on the distortion-rate characteristic of the optimal source coder for the input source of independent and identically distributed (i.i.d.) Gaussian random variables. We use the parametric methodology to derive some useful theoretical results, and then exploit those results to propose an optimization method where the joint assignment of spectral efficiencies (i.e., channel codes and modulations) and spatial multiplexing rates (i.e., space-time codes) to progressive packets can be optimized in a packet-by-packet manner. As a

result, the computational complexity involved in the optimization decreases exponentially relative to a conventional exhaustive search. Our work can be applied to the state-of-the-art wireless communication systems such as 3GPP Long Term Evolution and 5G systems based on orthogonal frequency division multiplexing (OFDM) for the delivery of progressive images. The rest of this paper is organized as follows. In Section II, we provide the system model and the technical preliminaries. In Section III, we propose the joint optimization method in a MIMO system. In Section IV, the performance of the proposed optimization method is numerically evaluated, and we conclude this paper in Section V.

II. Preliminaries

A. Progressive Systems

We describe the evaluation of the expected distortion of the progressive source. The system takes a compressed progressive bitstream from the source encoder, and transforms it into a sequence of packets with error detection capability. Then, as shown in Fig. 1, the packets are encoded using channel codes, modulations and space-time codes. At the receiver, if a packet has been correctly received, decoding of the next packet is considered by the source decoder. Otherwise, the decoding stops and the source is reconstructed from only the correctly decoded packets [2].

Let $p(u_i, v_i)$ denote the probability of the packet error with a spectral efficiency $u_i \in \mathcal{R} = \{R_1, R_2, \dots, R_{N_{\text{se}}}\}$, and spatial multiplexing rate $v_i \in \mathcal{C} = \{C_1, C_2, \dots, C_{N_{\text{smr}}}\}$, where N_{se} is the number of candidate spectral efficiencies, specified by channel codes and modulations, and N_{smr} is the number of candidate spatial multiplexing rates specified by the space-time codes employed in a system. We let $b(u_i, v_i)$ denote the number of information (or source) bits in the packet that employs $u_i \in \mathcal{R}$ and $v_i \in \mathcal{C}$. Recall that we have $b(u_i, v_i) = u_i v_i T_{\text{pkt}} W_{\text{pkt}}$, where T_{pkt} and W_{pkt} are the time duration and the signal bandwidth of the coded packet, respectively. Regarding $\mathcal{R}$ and $\mathcal{C}$, it is assumed that, for $u_i < u_j$ $(u_i, u_j \in \mathcal{R})$ and $v_i \in \mathcal{C}$, we have [8]

$$p(u_i, v_i) < p(u_j, v_i), \tag{1}$$

which indicates that, when the spatial multiplexing rate is fixed, a larger spectral efficiency (i.e., a larger channel code rate or a larger modulation alphabet size) leads to a higher probability of

error. It is also assumed that, for $v_i < v_j$ $(v_i, v_j \in \mathcal{C})$ and $u_i \in \mathcal{R}$, we have [16]

$$p(u_i, v_i) < p(u_i, v_j). \tag{2}$$

That is, for a fixed spectral efficiency (i.e., a fixed channel code rate and a fixed modulation alphabet size), a larger spatial multiplexing rate yields a higher probability of error.

Let $D_{1,2,\ldots,k}(u_1, u_2, \ldots, u_k;\, v_1, v_2, \ldots, v_k)$ be the expected distortion of the progressive source for the event where $u_i \in \mathcal{R}$ and $v_i \in \mathcal{C}$ are assigned to the ith packet ($i = 1, 2, \ldots, k$) in a sequence of k packets. We denote the operational distortion-rate function of the source by $f(x)$. From the aforementioned decoding rule of the progressive codes, $D_{1,2,\ldots,k}(u_1, u_2, \ldots, u_k;$ $v_1, v_2, \ldots, v_k)$ can be expressed as

$$D_{1,2,\ldots,k}(u_1, u_2, \ldots, u_k;\, v_1, v_2, \ldots, v_k) = \sum_{n=0}^{k} f\left(\sum_{i=1}^{n} b(u_i, v_i)\right) P_{\mathrm{c},\, n}, \tag{3}$$

where $f\big(\sum_{i=1}^{n} b(u_i, v_i)\big)$ is the distortion of the source for the case where the first n packets in a sequence of k packets are used for the source decoding, and $P_{\mathrm{c},\, n}$ is the probability that no decoding errors occur in the first n packets with an error in the next one. For $1 \leq n \leq k-1$, $P_{\mathrm{c},\, n}$ is given by $P_{\mathrm{c},\, n} = p(u_{n+1}, v_{n+1}) \prod_{i=1}^{n} \big(1 - p(u_i, v_i)\big)$; the probability of an error in the first packet is $P_{\mathrm{c},\, 0} = p(u_1, v_1)$, and the probability that all k packets are correctly decoded is $P_{\mathrm{c},\, k} = \prod_{i=1}^{k} \big(1 - p(u_i, v_i)\big)$. Then, (3) can be rewritten as

$$\begin{aligned} &D_{1,2,\ldots,k}(u_1, u_2, \ldots, u_k;\, v_1, v_2, \ldots, v_k) \\ &= \sum_{n=0}^{k-1} f\left(\sum_{i=1}^{n} b(u_i, v_i)\right) p(u_{n+1}, v_{n+1}) \prod_{i=1}^{n} \big(1 - p(u_i, v_i)\big) + f\left(\sum_{i=1}^{k} b(u_i, v_i)\right) \prod_{i=1}^{k} \big(1 - p(u_i, v_i)\big), \end{aligned} \tag{4}$$

where we have used the definitions of $\sum_{i=I_1}^{I_2} a(i) \triangleq 0$ and $\prod_{i=I_1}^{I_2} a(i) \triangleq 1$ for an arbitrary function $a(i)$.

B. MIMO Systems

We consider the transmission of progressive packets in a MIMO system with N_{t} transmit and N_{r} receive antennas. A space-time codeword, $\mathbf{S} = [\mathbf{s}_1\, \mathbf{s}_2\, \cdots\, \mathbf{s}_{T_{\mathrm{s}}}]$ of size $N_{\mathrm{t}} \times T_{\mathrm{s}}$ is transmitted over N_{t} transmit antennas and T_{s} symbol durations. After matched filtering and sampling, the $N_{\mathrm{r}} \times 1$ received signal vector, $\mathbf{y}_l$ $(1 \leq l \leq T_{\mathrm{s}})$, can be expressed as

$$\mathbf{y}_l = \mathbf{H}\mathbf{s}_l + \mathbf{n}_l, \tag{5}$$

where $\mathbf{s}_l$ is an $N_\mathrm{t} \times 1$ transmitted signal vector, and $\mathbf{n}_l$ is an $N_\mathrm{r} \times 1$ zero-mean complex additive white Gaussian noise (AWGN) vector with $\mathcal{E}\left[\mathbf{n}_k \mathbf{n}_l^H\right] = \sigma_\mathrm{n}^2 \mathbf{I}_{N_\mathrm{r}} \delta(k-l)$, where $(\cdot)^H$ denotes Hermitian operation, $\mathbf{I}_n$ is the $n \times n$ identity matrix, and $\delta(\cdot)$ denotes the Kronecker delta function. $\mathbf{H}$ denotes the $N_\mathrm{r} \times N_\mathrm{t}$ channel matrix, whose entry h_{ij} represents the complex channel coefficient between the jth transmit antenna and the ith receive antenna. The channel matrix can be modeled as [25]

$$\mathbf{H} = \sqrt{\frac{K}{K+1}}\bar{\mathbf{H}} + \sqrt{\frac{1}{K+1}}\mathbf{R}_\mathrm{r}^{1/2}\,\mathbf{H}_\mathbf{w}\,\mathbf{R}_\mathrm{t}^{1/2}, \tag{6}$$

where $K > 0$ is the Rician factor, and $\bar{\mathbf{H}}$ represents the channel matrix related to the line-of-sight (LOS) signal components. The Frobenius norm of $\bar{\mathbf{H}}$ is normalized as $(N_\mathrm{r} N_\mathrm{t})^{1/2}$, and $\bar{\mathbf{H}}$ is assumed to be known to both the transmitter and the receiver. $\mathbf{R}_\mathrm{t}$ is an $N_\mathrm{t} \times N_\mathrm{t}$ transmit spatial correlation matrix, $\mathbf{R}_\mathrm{r}$ is an $N_\mathrm{r} \times N_\mathrm{r}$ receive spatial correlation matrix, and $(\cdot)^{1/2}$ represents the Hermitian square root of a matrix. We use the exponential correlation model with $(\mathbf{R}_\mathrm{t})_{i,j} = \rho_\mathrm{t}^{|i-j|}$ and $(\mathbf{R}_\mathrm{r})_{i,j} = \rho_\mathrm{r}^{|i-j|}$, where $(\cdot)_{i,j}$ denotes the (i,j)th entry of a matrix, and ρ_t and ρ_r are the transmit and receive spatial correlation coefficients between adjacent antennas, respectively. In (6), $\mathbf{H}_\mathbf{w}$ is an $N_\mathrm{r} \times N_\mathrm{t}$ channel matrix whose entries are i.i.d. $\sim \mathcal{CN}(0,1)$, and $\mathbf{H}_\mathbf{w}$ is assumed to be known at the receiver, but not known at the transmitter (that is, the channel state information is available only at the receiver). It is assumed that $\mathbf{H}_\mathbf{w}$ is random, but constant over T_s symbol durations. Let γ_s denote the signal-to-noise ratio (SNR) per symbol, which is defined as $\gamma_\mathrm{s} := \mathcal{E}\left[|(\mathbf{s}_k)_i|^2\right]/\sigma_\mathrm{n}^2$ where $(\cdot)_i$ denotes the ith component of a vector. Let N_s denote the number of symbols packed within a space-time codeword $\mathbf{S} = [\mathbf{s}_1\,\mathbf{s}_2\,\cdots\,\mathbf{s}_{T_\mathrm{s}}]$. Then, the spatial multiplexing rate of $\mathbf{S}$ is given by $N_\mathrm{s}/T_\mathrm{s}$.

III. Optimization of Joint Progressive Source, Channel, and Space-time Coding

The joint optimization problem for a sequence of k progressive packets in a MIMO system is to find the set of spectral efficiencies $u_1, u_2, \ldots, u_k \in \mathcal{R} = \{R_1, R_2, \ldots, R_{N_\mathrm{se}}\}$ and spatial multiplexing rates $v_1, v_2, \ldots, v_k \in \mathcal{C} = \{C_1, C_2, \ldots, C_{N_\mathrm{smr}}\}$ that minimizes the expected distortion given by (4). The number of possible assignments of N_se spectral efficiencies and N_smr spatial multiplexing rates to k packets is $(N_\mathrm{se} N_\mathrm{smr})^k$, which exponentially grows as k increases. Thus, when k is large, an exhaustive search of the optimal solution would be computationally prohibitive. In addition, due to the high nonlinearity of $f(x)$ and $p(x,y)$ in the expected distortion

given by (4), convex optimization techniques are not exploited to find the optimal solution. To address this matter, we consider a parametric model of the operational distortion-rate characteristic of the source. When a sequence of i.i.d. Gaussian random variables with zero mean and variance of σ^2 are encoded at bit rate x using an optimal source coder, the distortion of the source is given by [26]

$$d(x) = \sigma^2 2^{-2x}. \tag{7}$$

Although the distortion-rate bound, given by (7), can only be achieved with no constraint on the coding length, the operational distortion-rate function of a practical source coder usually shows the same exponential rate decay of $2x$ at a high bit rate [16].

Instead of the actual operational distortion-rate function of the source, denoted by $f(x)$, we take into account the parametric distortion-rate function:

$$f^{\mathrm{p}}(x) = \sigma^2 2^{-\alpha x}, \quad \alpha \geq 2 \tag{8}$$

where α is a parameter that is free to be adjusted for the optimization, and σ^2 $(\neq 0)$ is a constant that does not affect the optimization (this will be described in detail in this section). The function in (8) parameterizes the distortion-rate characteristic of the source $d(x)$, given by (7), to include a wide range of low-to-high bit rates x for a practical image source coder. Suppose that instead of $f(x)$, $f^{\mathrm{p}}(x)$ is employed for the computation of the expected distortion. Then, from (4), the resulting expected distortion, denoted by $D^{\mathrm{p}}_{1,2,\ldots,k}(u_1, u_2, \ldots, u_k;\ v_1, v_2, \ldots, v_k;\ \alpha)$, can be expressed as

$$\begin{aligned} &D^{\mathrm{p}}_{1,2,\ldots,k}(u_1, u_2, \ldots, u_k;\ v_1, v_2, \ldots, v_k;\ \alpha) \\ &\quad = \sum_{n=0}^{k-1} \sigma^2 \Bigg(\prod_{i=1}^{n} g\big(b(u_i, v_i)\big) \Bigg) p(u_{n+1}, v_{n+1}) \prod_{i=1}^{n} \big(1 - p(u_i, v_i)\big) \\ &\quad + \sigma^2 \prod_{i=1}^{k} g\big(b(u_i, v_i)\big) \prod_{i=1}^{k} \big(1 - p(u_i, v_i)\big), \end{aligned} \tag{9}$$

where $g(x) \triangleq 2^{-\alpha x}$. Let $\mathbf{s}_{1,2,\ldots,k} = [u_1, u_2, \ldots, u_k;\ v_1, v_2, \ldots, v_k]$ indicate a solution (or assignment) where a spectral efficiency $u_i \in \mathcal{R}$ and a spatial multiplexing rate $v_i \in \mathcal{C}$ are assigned to the ith packet ($i = 1, 2, \ldots, k$). From here onwards, we refer to the parametric distortion-based optimal solution as the one that minimizes the expected distortion of the source with the parametric distortion-rate function $f^{\mathrm{p}}(x)$ employed as given by (9). In the following, we derive

some theoretical results for the parametric distortion-based optimal solution. Based on them, we will propose an efficient optimization method for the joint source, channel, and space-time coding of progressive packets.

Theorem 1: For some integer L in the range of $L \geq 2$, suppose that $\mathbf{s}_{1,2,\ldots,L-1} = [r_2^*,\, r_3^*,\, \ldots,\, r_L^*;\, c_2^*,\, c_3^*,\, \ldots,\, c_L^*]$ is a parametric distortion-based optimal assignment of spectral efficiencies and spatial multiplexing rates to $L-1$ progressive packets. That is,

$$D^{\mathrm{p}}_{1,2,\ldots,L-1}(r_2^*, r_3^*, \ldots, r_L^*;\ c_2^*, c_3^*, \ldots, c_L^*;\ \alpha) \leq D^{\mathrm{p}}_{1,2,\ldots,L-1}(r_2, r_3, \ldots, r_L;\ c_2, c_3, \ldots, c_L;\ \alpha)$$
$$\text{for any } r_2, r_3, \ldots, r_L \in \mathcal{R} \text{ and } c_2, c_3, \ldots, c_L \in \mathcal{C}. \tag{10}$$

Then, if

$$D^{\mathrm{p}}_{1,2,\ldots,L}(r_1^*, r_2^*, \ldots, r_L^*;\ c_1^*, c_2^*, \ldots, c_L^*;\ \alpha) \leq D^{\mathrm{p}}_{1,2,\ldots,L}(r_1, r_2^*, \ldots, r_L^*;\ c_1, c_2^*, \ldots, c_L^*;\ \alpha)$$
$$\text{for any } r_1 \in \mathcal{R} \text{ and } c_1 \in \mathcal{C}, \tag{11}$$

we obtain

$$D^{\mathrm{p}}_{1,2,\ldots,L}(r_1^*, r_2^*, \ldots, r_L^*;\ c_1^*, c_2^*, \ldots, c_L^*;\ \alpha) \leq D^{\mathrm{p}}_{1,2,\ldots,L}(r_1, r_2, \ldots, r_L;\ c_1, c_2, \ldots, c_L;\ \alpha)$$
$$\text{for any } r_1, r_2, \ldots, r_L \in \mathcal{R} \text{ and } c_1, c_2, \ldots, c_L \in \mathcal{C}. \tag{12}$$

In other words, $\mathbf{s}_{1,2,\ldots,L} = [r_1^*, r_2^*, \ldots, r_L^*;\ c_1^*, c_2^*, \ldots, c_L^*]$ is a parametric distortion-based optimal solution for L progressive packets.

Proof: Consider the case where spectral efficiencies $r_1, r_2, \ldots, r_L$ and spatial multiplexing rates $c_1, c_2, \ldots, c_L$ are assigned to packets $1, 2, \ldots, L$, respectively. From (9), it can be shown that the corresponding expected distortion, $D^{\mathrm{p}}_{1,2,\ldots,L}(r_1, r_2, \ldots, r_L;\, c_1, c_2^*, \ldots, c_L;\, \alpha)$, is expressed as

$$\begin{aligned} D^{\mathrm{p}}_{1,2,\ldots,L}(r_1, r_2, \ldots, r_L;\ c_1, c_2, \ldots, c_L;\ \alpha) &= \sigma^2 p(r_1, c_1) + \sigma^2 g\big(b(r_1, c_1)\big)\big(1 - p(r_1, c_1)\big) \\ \times \Bigg[\sum_{n=1}^{L-1} \Bigg(\prod_{i=2}^{n} g\big(b(r_i, c_i)\big) \Bigg) p(r_{n+1}, c_{n+1}) \prod_{i=2}^{n} \big(1 - p(r_i, c_i)\big) &+ \prod_{i=2}^{L} g\big(b(r_i, c_i)\big) \prod_{i=2}^{L} \big(1 - p(r_i, c_i)\big) \Bigg]. \end{aligned} \tag{13}$$

Consider another case where spectral efficiencies $r_2, r_3, \ldots, r_L$ and spatial multiplexing rates $c_2, c_3, \ldots, c_L$ are assigned to packets $1, 2, \ldots, L-1$, respectively. From (9), it can be shown that

the resulting expected distortion, $D^{\mathrm{p}}_{1,2,\ldots,L-1}(r_2, r_3, \ldots, r_L;\ c_2, c_3, \ldots, c_L;\ \alpha)$, is given by

$$D^{\mathrm{p}}_{1,2,\ldots,L-1}(r_2, r_3, \ldots, r_L;\ c_2, c_3, \ldots, c_L;\ \alpha) = \sigma^2 \sum_{n=1}^{L-1} \left(\prod_{i=2}^{n} g\big(b(r_i, c_i)\big) \right) p(r_{n+1}, c_{n+1}) \times \prod_{i=2}^{n} \big(1 - p(r_i, c_i)\big) + \sigma^2 \prod_{i=2}^{L} g\big(b(r_i, c_i)\big) \prod_{i=2}^{L} \big(1 - p(r_i, c_i)\big). \quad (14)$$

From (13) and (14), we have

$$D^{\mathrm{p}}_{1,2,\ldots,L}(r_1, r_2, \ldots, r_L;\ c_1, c_2, \ldots, c_L;\ \alpha) = \sigma^2 p(r_1, c_1) + g\big(b(r_1, c_1)\big)\big(1 - p(r_1, c_1)\big) \times D^{\mathrm{p}}_{1,2,\ldots,L-1}(r_2, r_3, \ldots, r_L;\ c_2, c_3, \ldots, c_L;\ \alpha). \quad (15)$$

Then, the inequality given by (11) can be rewritten as

$$\begin{aligned} & D^{\mathrm{p}}_{1,2,\ldots,L}(r_1^*, r_2^*, \ldots, r_L^*;\ c_1^*, c_2^*, \ldots, c_L^*;\ \alpha) \\ & \le \sigma^2 p(r_1, c_1) + g\big(b(r_1, c_1)\big)\big(1 - p(r_1, c_1)\big) D^{\mathrm{p}}_{1,2,\ldots,L-1}(r_2^*, r_3^*, \ldots, r_L^*;\ c_2^*, c_3^*, \ldots, c_L^*;\ \alpha) \\ & \le \sigma^2 p(r_1, c_1) + g\big(b(r_1, c_1)\big)\big(1 - p(r_1, c_1)\big) D^{\mathrm{p}}_{1,2,\ldots,L-1}(r_2, r_3, \ldots, r_L;\ c_2, c_3, \ldots, c_L;\ \alpha) \\ & = D^{\mathrm{p}}_{1,2,\ldots,L}(r_1, r_2, \ldots, r_L;\ c_1, c_2, \ldots, c_L;\ \alpha) \\ & \qquad \text{for any } r_1, r_2, \ldots, r_L \in \mathcal{R} \text{ and } c_1, c_2, \ldots, c_L \in \mathcal{C}. \end{aligned} \quad (16)$$

where both the first inequality and the equality follow from (15), and the second inequality follows from the supposition of (10). Thus, (12) is valid.

□

Theorem 1 tells us that if the parametric distortion-rate function $f^{\mathrm{p}}(x)$, given by (8), is used to compute the expected distortion, the joint assignment of the spectral efficiencies and spatial multiplexing rates to progressive packets can be optimized in a packet-by-packet manner. This optimization method will be described in detail in this section.

Lemma 2: Consider some integer i in the range of $1 \le i \le L-1$. If $\mathbf{s}_{1,2,\ldots,L-i+1} = [r_i^*, r_{i+1}^*, \ldots, r_L^*;\ c_i^*, c_{i+1}^*, \ldots, c_L^*]$ is a parametric distortion-based optimal assignment of spectral efficiencies and spatial multiplexing rates to $L-i+1$ progressive packets, i.e.,

$$\begin{aligned} & D^{\mathrm{p}}_{1,2,\ldots,L-i+1}(r_i^*, r_{i+1}^*, \ldots, r_L^*;\ c_i^*, c_{i+1}^*, \ldots, c_L^*;\ \alpha) \\ & \le D^{\mathrm{p}}_{1,2,\ldots,L-i+1}(r_i, r_{i+1}, \ldots, r_L;\ c_i, c_{i+1}, \ldots, c_L;\ \alpha) \\ & \text{for any } r_i, r_{i+1}, \ldots, r_L \in \mathcal{R} \text{ and } c_i, c_{i+1}, \ldots, c_L \in \mathcal{C}, \end{aligned} \quad (17)$$

then, for some integer j in the range of $i+1 \leq j \leq L$, we have

$$\begin{aligned}&D^{\mathrm{p}}_{1,2,\ldots,L-j+1}(r^*_j, r^*_{j+1}, \ldots, r^*_L;\ c^*_j, c^*_{j+1}, \ldots, c^*_L;\ \alpha)\\&\leq D^{\mathrm{p}}_{1,2,\ldots,L-j+1}(r_j, r_{j+1}, \ldots, r_L;\ c_j, c_{j+1}, \ldots, c_L;\ \alpha)\\&\text{for any } r_j, r_{j+1}, \ldots, r_L \in \mathcal{R} \text{ and } c_j, c_{j+1}, \ldots, c_L \in \mathcal{C}.\end{aligned} \tag{18}$$

In other words, $\mathbf{s}_{1,2,\ldots,L-j+1} = [r^*_j, r^*_{j+1}, \ldots, r^*_L;\ c^*_j, c^*_{j+1}, \ldots, c^*_L]$ is a parametric distortion-based optimum for $L-j+1$ progressive packets.

Proof: From (17), it follows that

$$\begin{aligned}&D^{\mathrm{p}}_{1,2,\ldots,L-i+1}(r^*_i, r^*_{i+1}, \ldots, r^*_L;\ c^*_i, c^*_{i+1}, \ldots, c^*_L;\ \alpha)\\&\leq D^{\mathrm{p}}_{1,2,\ldots,L-i+1}(r^*_i, r_{i+1}, \ldots, r_L;\ c^*_i, c_{i+1}, \ldots, c_L;\ \alpha)\\&\text{for any } r_{i+1}, r_{i+2}, \ldots, r_L \in \mathcal{R} \text{ and } c_{i+1}, c_{i+2}, \ldots, c_L \in \mathcal{C}.\end{aligned} \tag{19}$$

From (15), the inequality given by (19) can be rewritten as

$$\begin{aligned}&\sigma^2 p(r^*_i, c^*_i) + g\big(b(r^*_i, c^*_i)\big)\big(1 - p(r^*_i, c^*_i)\big) D^{\mathrm{p}}_{1,2,\ldots,L-i}(r^*_{i+1}, r^*_{i+2}, \ldots, r^*_L;\ c^*_{i+1}, c^*_{i+2}, \ldots, c^*_L;\ \alpha)\\&\leq \sigma^2 p(r^*_i, c^*_i) + g\big(b(r^*_i, c^*_i)\big)\big(1 - p(r^*_i, c^*_i)\big) D^{\mathrm{p}}_{1,2,\ldots,L-i}(r_{i+1}, r_{i+2}, \ldots, r_L;\ c_{i+1}, c_{i+2}, \ldots, c_L;\ \alpha)\\&\text{for any } r_{i+1}, r_{i+2}, \ldots, r_L \in \mathcal{R} \text{ and } c_{i+1}, c_{i+2}, \ldots, c_L \in \mathcal{C}.\end{aligned} \tag{20}$$

Since $p(r^*_i, c^*_i) < 1$, $g\big(b(r^*_i, c^*_i)\big) = 2^{-\alpha r^*_i c^*_i T_{\mathrm{pkt}} W_{\mathrm{pkt}}} > 0$, and from (20), we have

$$\begin{aligned}&D^{\mathrm{p}}_{1,2,\ldots,L-i}(r^*_{i+1}, r^*_{i+2}, \ldots, r^*_L;\ c^*_{i+1}, c^*_{i+2}, \ldots, c^*_L;\ \alpha)\\&\leq D^{\mathrm{p}}_{1,2,\ldots,L-i}(r_{i+1}, r_{i+2}, \ldots, r_L;\ c_{i+1}, c_{i+2}, \ldots, c_L;\ \alpha)\\&\text{for any } r_{i+1}, r_{i+2}, \ldots, r_L \in \mathcal{R} \text{ and } c_{i+1}, c_{i+2}, \ldots, c_L \in \mathcal{C}.\end{aligned} \tag{21}$$

Based on (21), we will prove (18) by induction on the number of packets: We first consider $L-i$ packets. Eq. (21) is identical to (18) when we let $j = i+1$ in (18) (i.e., $L-i$ packets). We next suppose that (18) holds for $j = n$ $(\geq i+1)$. In other words, for $L-n+1$ ($\leq L-i$) packets, we have

$$\begin{aligned}&D^{\mathrm{p}}_{1,2,\ldots,L-n+1}(r^*_n, r^*_{n+1}, \ldots, r^*_L;\ c^*_n, c^*_{n+1}, \ldots, c^*_L;\ \alpha)\\&\leq D^{\mathrm{p}}_{1,2,\ldots,L-n+1}(r_n, r_{n+1}, \ldots, r_L;\ c_n, c_{n+1}, \ldots, c_L;\ \alpha)\\&\text{for any } r_n, r_{n+1}, \ldots, r_L \in \mathcal{R} \text{ and } c_n, c_{n+1}, \ldots, c_L \in \mathcal{C}.\end{aligned} \tag{22}$$

Eq. (22), which is the induction hypothesis, is identical to (17) when we let $i = n$ in (17). Since (17) implies (21), (21) holds for $i = n$, i.e.,

$$\begin{aligned}
&D^{\mathrm{p}}_{1,2,\ldots,L-n}(r^*_{n+1}, r^*_{n+2}, \ldots, r^*_L;\ c^*_{n+1}, c^*_{n+2}, \ldots, c^*_L;\ \alpha) \\
&\leq D^{\mathrm{p}}_{1,2,\ldots,L-n}(r_{n+1}, r_{n+2}, \ldots, r_L;\ c_{n+1}, c_{n+2}, \ldots, c_L;\ \alpha) \\
&\text{for any } r_{n+1}, r_{n+2}, \ldots, r_L \in \mathcal{R} \text{ and } c_{n+1}, c_{n+2}, \ldots, c_L \in \mathcal{C}.
\end{aligned} \tag{23}$$

Letting $j = n+1$ in (18) (i.e., $L-n$ packets), we obtain a result that is identical to (23). Hence, (18) holds for $j = n+1$. We have thus shown that (18) is valid for $j \geq i+1$.

□

Lemma 2 tells us that if a parametric distortion-based optimal solution for a given number of packets (or, equivalently, a given transmission rate in bpp) has been obtained, an optimal solution for a smaller number of packets (or a lower transmission rate) can be immediately found without additional computation. As an example, if $\mathbf{s}_{1,2,3} = [r^*_1, r^*_2, r^*_3;\ c^*_1, c^*_2, c^*_3]$ is a parametric distortion-based optimum for a sequence of three packets ($L = 3$), then $\mathbf{s}_{1,2} = [r^*_2, r^*_3;\ c^*_2, c^*_3]$ is an optimum for a sequence of two packets ($L = 2$), and $\mathbf{s}_1 = [r^*_3;\ c^*_3]$ is an optimum for a single packet ($L = 1$). Lemma 2 is used to prove the subsequent Lemma 4, Theorem 5, and Corollary 6.

Lemma 3: For some integer L in the range of $L \geq 2$, we have

$$\begin{aligned}
&D^{\mathrm{p}}_{1,2,\ldots,L}(r_1, r_2, \ldots, r_L;\ c_1, c_2, \ldots, c_L;\ \alpha) < D^{\mathrm{p}}_{1,2,\ldots,L-1}(r_1, r_2, \ldots, r_{L-1};\ c_1, c_2, \ldots, c_{L-1};\ \alpha) \\
&\qquad\text{for any } r_1, r_2, \ldots, r_L \in \mathcal{R} \text{ and } c_1, c_2, \ldots, c_L \in \mathcal{C}.
\end{aligned} \tag{24}$$

Proof: From (9), it can be shown that $D^{\mathrm{p}}_{1,2,\ldots,L}(r_1, r_2, \ldots, r_L;\ c_1, c_2, \ldots, c_L;\ \alpha)$ is rewritten as

$$\begin{aligned}
&D^{\mathrm{p}}_{1,2,\ldots,L}(r_1, r_2, \ldots, r_L;\ c_1, c_2, \ldots, c_L;\ \alpha) \\
&= \sum_{n=0}^{L-2} \sigma^2 \left(\prod_{i=1}^{n} g\big(b(r_i, c_i)\big) \right) p(r_{n+1}, c_{n+1}) \prod_{i=1}^{n} \big(1 - p(r_i, c_i)\big) \\
&+ \sigma^2 \prod_{i=1}^{L-1} g\big(b(r_i, c_i)\big) \prod_{i=1}^{L-1} \big(1 - p(r_i, c_i)\big) \Big\{ p(r_L, c_L) + g\big(b(r_L, c_L)\big)\big(1 - p(r_L, c_L)\big) \Big\}.
\end{aligned} \tag{25}$$

In addition, from (9), $D^{\mathrm{p}}_{1,2,\ldots,L-1}(r_1, r_2, \ldots, r_{L-1};\ c_1, c_2, \ldots, c_{L-1};\ \alpha)$ is given by

$$\begin{aligned} &D^{\mathrm{p}}_{1,2,\ldots,L-1}(r_1, r_2, \ldots, r_{L-1};\ c_1, c_2, \ldots, c_{L-1};\ \alpha) \\ &\quad = \sum_{n=0}^{L-2} \sigma^2 \left(\prod_{i=1}^{n} g\big(b(r_i, c_i)\big) \right) p(r_{n+1}, c_{n+1}) \prod_{i=1}^{n} \big(1 - p(r_i, c_i)\big) \\ &\quad + \sigma^2 \prod_{i=1}^{L-1} g\big(b(r_i, c_i)\big) \prod_{i=1}^{L-1} \big(1 - p(r_i, c_i)\big). \end{aligned} \tag{26}$$

From (25) and (26), we obtain

$$\begin{aligned} &D^{\mathrm{p}}_{1,2,\ldots,L}(r_1, r_2, \ldots, r_L;\ c_1, c_2, \ldots, c_L;\ \alpha) \\ &\quad = D^{\mathrm{p}}_{1,2,\ldots,L-1}(r_1, r_2, \ldots, r_{L-1};\ c_1, c_2, \ldots, c_{L-1};\ \alpha) \\ &\qquad + \sigma^2 \prod_{i=1}^{L-1} g\big(b(r_i, c_i)\big) \prod_{i=1}^{L-1} \big(1 - p(r_i, c_i)\big) \Big\{ p(r_L, c_L) + g\big(b(r_L, c_L)\big)\big(1 - p(r_L, c_L)\big) - 1 \Big\} \\ &\quad < D^{\mathrm{p}}_{1,2,\ldots,L-1}(r_1, r_2, \ldots, r_{L-1};\ c_1, c_2, \ldots, c_{L-1};\ \alpha), \end{aligned} \tag{27}$$

where the inequality follows from $p(r_L, c_L) < 1$, and $0 < g\big(b(r_L, c_L)\big) = 2^{-\alpha r_L c_L T_{\mathrm{pkt}} W_{\mathrm{pkt}}} < 1$.

$\square$

Lemma 3 is used in the proof of Lemma 4.

Lemma 4: Consider some integers i, j in the range of $1 \leq i \leq L-1$ and $i+1 \leq j \leq L$. If $\mathbf{s}_{1,2,\ldots,L-i+1} = [r_i^*, r_{i+1}^*, \ldots, r_L^*;\ c_i^*, c_{i+1}^*, \ldots, c_L^*]$ is a parametric distortion-based optimum for $L-i+1$ progressive packets (that is, if (17) holds), we obtain

$$\begin{aligned} &D^{\mathrm{p}}_{1,2,\ldots,L-i+1}(r_i^*, r_{i+1}^*, \ldots, r_L^*;\ c_i^*, c_{i+1}^*, \ldots, c_L^*;\ \alpha) \\ &< D^{\mathrm{p}}_{1,2,\ldots,L-j+1}(r_j^*, r_{j+1}^*, \ldots, r_L^*;\ c_j^*, c_{j+1}^*, \ldots, c_L^*;\ \alpha). \end{aligned} \tag{28}$$

Proof: From (24) of Lemma 3, it is clear that for $1 \leq i \leq L-1$, we have

$$\begin{aligned} &D^{\mathrm{p}}_{1,2,\ldots,L-i+1}(r_{i+1}^*, r_{i+2}^*, \ldots, r_L^*, r_k;\ c_{i+1}^*, c_{i+2}^*, \ldots, c_L^*, c_k;\ \alpha) \\ &< D^{\mathrm{p}}_{1,2,\ldots,L-i}(r_{i+1}^*, r_{i+2}^*, \ldots, r_L^*;\ c_{i+1}^*, c_{i+2}^*, \ldots, c_L^*;\ \alpha) \\ &\qquad \text{for any } r_k \in \mathcal{R} \text{ and } c_k \in \mathcal{C}. \end{aligned} \tag{29}$$

From the condition of this lemma given by (17) and (29), we can derive

$$\begin{aligned}
&D^{\mathrm{p}}_{1,2,\ldots,L-i+1}(r^*_i, r^*_{i+1}, \ldots, r^*_L;\; c^*_i, c^*_{i+1}, \ldots, c^*_L;\; \alpha)\\
&\leq D^{\mathrm{p}}_{1,2,\ldots,L-i+1}(r^*_{i+1}, r^*_{i+2}, \ldots, r^*_L, r_k;\; c^*_{i+1}, c^*_{i+2}, \ldots, c^*_L, c_k;\; \alpha)\\
&< D^{\mathrm{p}}_{1,2,\ldots,L-i}(r^*_{i+1}, r^*_{i+2}, \ldots, r^*_L;\; c^*_{i+1}, c^*_{i+2}, \ldots, c^*_L;\; \alpha)\\
&\qquad\qquad \text{for any } r_k \in \mathcal{R} \text{ and } c_k \in \mathcal{C},
\end{aligned} \tag{30}$$

where the first inequality follows from (17), and the second inequality follows from (29).

Based on (30), we will prove (28) by induction on the number of packets: We first consider $L-i$ packets. Eq. (30) is identical to (28) when we let $j = i+1$ in (28) (i.e., $L-i$ packets). We next suppose that (28) holds for $j = n$ $(\geq i+1)$. In other words, for $L-n+1$ $(\leq L-i)$ packets, we have the following induction hypothesis.

$$\begin{aligned}
&D^{\mathrm{p}}_{1,2,\ldots,L-i+1}(r^*_i, r^*_{i+1}, \ldots, r^*_L;\; c^*_i, c^*_{i+1}, \ldots, c^*_L;\; \alpha)\\
&< D^{\mathrm{p}}_{1,2,\ldots,L-n+1}(r^*_n, r^*_{n+1}, \ldots, r^*_L;\; c^*_n, c^*_{n+1}, \ldots, c^*_L;\; \alpha).
\end{aligned} \tag{31}$$

Note that the right hand side of (31) is also a parametric distortion-based optimum for $L-n+1$ progressive packets, because Lemma 2 indicates that the condition of this lemma, which is given by (17), implies (18) for some integer $j \geq i+1$. From the fact that a parametric distortion-based optimal solution satisfies (30), and that the right hand side of (31) equals the first line of (30) when setting $i = n$ in (30), it follows that

$$\begin{aligned}
&D^{\mathrm{p}}_{1,2,\ldots,L-n+1}(r^*_n, r^*_{n+1}, \ldots, r^*_L;\; c^*_n, c^*_{n+1}, \ldots, c^*_L;\; \alpha)\\
&< D^{\mathrm{p}}_{1,2,\ldots,L-n}(r^*_{n+1}, r^*_{n+2}, \ldots, r^*_L;\; c^*_{n+1}, c^*_{n+2}, \ldots, c^*_L;\; \alpha).
\end{aligned} \tag{32}$$

From the induction hypothesis given by (31) and (32), we have

$$\begin{aligned}
&D^{\mathrm{p}}_{1,2,\ldots,L-i+1}(r^*_i, r^*_{i+1}, \ldots, r^*_L;\; c^*_i, c^*_{i+1}, \ldots, c^*_L;\; \alpha)\\
&< D^{\mathrm{p}}_{1,2,\ldots,L-n}(r^*_{n+1}, r^*_{n+2}, \ldots, r^*_L;\; c^*_{n+1}, c^*_{n+2}, \ldots, c^*_L;\; \alpha).
\end{aligned} \tag{33}$$

Letting $j = n+1$ in (28) (i.e., $L-n$ packets), we obtain a result identical to (33). Hence, (28) holds for $j = n+1$. We have thus shown that (28) holds for $j \geq i+1$.

□

Lemma 4 is employed in the proof of Theorem 5, which derives some constraints on the search space of $\mathcal{R}$ and $\mathcal{C}$ when we find a parametric distortion-based optimal solution.

Theorem 5: Consider some integer i in the range of $1 \leq i \leq L-1$. If $\mathbf{s}_{1,2,\ldots,L-i+1} = [r_i^*, r_{i+1}^*, \ldots, r_L^*; c_i^*, c_{i+1}^*, \ldots, c_L^*]$ is a parametric distortion-based optimum for $L-i+1$ progressive packets (that is, if (17) holds), then for every integer j in the range of $i+1 \leq j \leq L$, at least one of the following three conditions holds:

$$\text{i) } r_i^* < r_j^*, \quad \text{ii) } c_i^* < c_j^*, \quad \text{iii) } r_i^* = r_j^*,\ c_i^* = c_j^*. \tag{34}$$

That is, we obtain at least $L-i$ constraints on r_i^* or c_i^* of the ith packet.

Proof: From the condition for this theorem, given by (17), for some integer j in the range of $i+1 \leq j \leq L$, we obtain

$$\begin{aligned} &D^{\mathrm{p}}_{1,2,\ldots,L-i+1}(r_i^*, r_{i+1}^*, \ldots, r_L^*;\ c_i^*, c_{i+1}^*, \ldots, c_L^*;\ \alpha) \\ &\leq D^{\mathrm{p}}_{1,2,\ldots,L-i+1}(r_j^*, r_{i+1}^*, r_{i+2}^*, \ldots, r_L^*;\ c_j^*, c_{i+1}^*, c_{i+2}^*, \ldots, c_L^*;\ \alpha). \end{aligned} \tag{35}$$

From (15), for $1 \leq i \leq L-1$ and $i+1 \leq j \leq L$, (35) can be rewritten as

$$\begin{aligned} \sigma^2\big(p(r_i^*, c_i^*) - p(r_j^*, c_j^*)\big) \leq \Big\{ & g\big(b(r_j^*, c_j^*)\big)\big(1 - p(r_j^*, c_j^*)\big) \\ & - g\big(b(r_i^*, c_i^*)\big)\big(1 - p(r_i^*, c_i^*)\big)\Big\} D^{\mathrm{p}}_{1,2,\ldots,L-i}(r_{i+1}^*, r_{i+2}^*, \ldots, r_L^*;\ c_{i+1}^*, c_{i+2}^*, \ldots, c_L^*;\ \alpha). \end{aligned} \tag{36}$$

Lemma 2 indicates that the condition of this theorem, given by (17), implies (18) for some integer j in the range of $i+1 \leq j \leq L$. Based on this, in the following, we will derive the inequality: For $1 \leq i \leq L-1$ and $i+1 \leq j \leq L$, we have

$$g\big(b(r_j^*, c_j^*)\big)\big(1 - p(r_j^*, c_j^*)\big) \leq g\big(b(r_i^*, c_i^*)\big)\big(1 - p(r_i^*, c_i^*)\big). \tag{37}$$

i) The case where j is in the range of $i+1 \leq j \leq L-1$

From (18), for $1 \leq i \leq L-2$ and $i+1 \leq j \leq L-1$, we obtain

$$\begin{aligned} &D^{\mathrm{p}}_{1,2,\ldots,L-j+1}(r_j^*, r_{j+1}^*, \ldots, r_L^*;\ c_j^*, c_{j+1}^*, \ldots, c_L^*;\ \alpha) \\ &\leq D^{\mathrm{p}}_{1,2,\ldots,L-j+1}(r_i^*, r_{j+1}^*, r_{j+2}^*, \ldots, r_L^*;\ c_i^*, c_{j+1}^*, c_{j+2}^*, \ldots, c_L^*;\ \alpha). \end{aligned} \tag{38}$$

Using (15), the inequality given by (38) can be rewritten as

$$\begin{aligned} \sigma^2\big(p(r_i^*, c_i^*) - p(r_j^*, c_j^*)\big) \geq \Big\{ & g\big(b(r_j^*, c_j^*)\big)\big(1 - p(r_j^*, c_j^*)\big) - g\big(b(r_i^*, c_i^*)\big) \\ & \times \big(1 - p(r_i^*, c_i^*)\big)\Big\} D^{\mathrm{p}}_{1,2,\ldots,L-j}(r_{j+1}^*, r_{j+2}^*, \ldots, r_L^*;\ c_{j+1}^*, c_{j+2}^*, \ldots, c_L^*;\ \alpha). \end{aligned} \tag{39}$$

From (36) and (39), it follows that for $1 \le i \le L-2$ and $i+1 \le j \le L-1$, we have

$$\begin{aligned}
&\left\{g\big(b(r_j^*, c_j^*)\big)\big(1-p(r_j^*, c_j^*)\big) - g\big(b(r_i^*, c_i^*)\big)\big(1-p(r_i^*, c_i^*)\big)\right\} \\
&\quad \times \left\{D^{\mathrm{p}}_{1,2,\ldots,L-i}(r_{i+1}^*, r_{i+2}^*, \ldots, r_L^*;\ c_{i+1}^*, c_{i+2}^*, \ldots, c_L^*;\ \alpha)\right. \\
&\quad\quad \left. - D^{\mathrm{p}}_{1,2,\ldots,L-j}(r_{j+1}^*, r_{j+2}^*, \ldots, r_L^*;\ c_{j+1}^*, c_{j+2}^*, \ldots, c_L^*;\ \alpha)\right\} \ge 0, \qquad (40)
\end{aligned}$$

where the ranges of i, j in (40) are subsets of those in (36) and (39). Lemma 4 shows that the condition of this theorem, given by (17), implies (28). If we let $i = k+1$ and $j = l+1$ in (28), then for $0 \le k \le L-2$ and $k+1 \le l \le L-1$, we have

$$\begin{aligned}
&D^{\mathrm{p}}_{1,2,\ldots,L-k}(r_{k+1}^*, r_{k+2}^*, \ldots, r_L^*;\ c_{k+1}^*, c_{k+2}^*, \ldots, c_L^*;\ \alpha) \\
&< D^{\mathrm{p}}_{1,2,\ldots,L-l}(r_{l+1}^*, r_{l+2}^*, \ldots, r_L^*;\ c_{l+1}^*, c_{l+2}^*, \ldots, c_L^*;\ \alpha), \qquad (41)
\end{aligned}$$

where the ranges of k and l are respectively derived from $1 \le i \le L-1$ and $i+1 \le j \le L$ in (28). From (40) and (41), it follows that (37) holds for $1 \le i \le L-2$ and $i+1 \le j \le L-1$.

ii) The case of $j = L$

From (18), for $1 \le i \le L-1$ and $j = L$ $(\ge i+1)$, we obtain

$$D^{\mathrm{p}}_1(r_L^*;\ c_L^*;\ \alpha) \le D^{\mathrm{p}}_1(r_i^*;\ c_i^*;\ \alpha). \qquad (42)$$

From the definitions below (4), letting $k = 1$ in (9) yields

$$D^{\mathrm{p}}_1(u_1;\ v_1;\ \alpha) = \sigma^2 p(u_1, v_1) + \sigma^2 g\big(b(u_1, v_1)\big)\big(1 - p(u_1, v_1)\big). \qquad (43)$$

Using (43), we can rewrite the inequality given by (42) as

$$p(r_i^*, c_i^*) - p(r_L^*, c_L^*) \ge g\big(b(r_L^*, c_L^*)\big)\big(1-p(r_L^*, c_L^*)\big) - g\big(b(r_i^*, c_i^*)\big)\big(1-p(r_i^*, c_i^*)\big). \qquad (44)$$

From (36) and (44), it follows that for $1 \le i \le L-1$ and $j = L$, we have

$$\begin{aligned}
&\left\{g\big(b(r_L^*, c_L^*)\big)\big(1-p(r_L^*, c_L^*)\big) - g\big(b(r_i^*, c_i^*)\big)\big(1-p(r_i^*, c_i^*)\big)\right\} \\
&\quad \times \left\{D^{\mathrm{p}}_{1,2,\ldots,L-i}(r_{i+1}^*, r_{i+2}^*, \ldots, r_L^*;\ c_{i+1}^*, c_{i+2}^*, \ldots, c_L^*;\ \alpha) - \sigma^2\right\} \ge 0, \qquad (45)
\end{aligned}$$

where the ranges of i, j in (45) are contained within those given by (36) and (44). In the following, we will show from (45) that (37) holds for $1 \le i \le L-1$ and $j = L$.

We first consider the case of $1 \leq i \leq L-2$ and $j = L$. Recall that the condition given by (17) for this theorem implies (28) of Lemma 4. If we set $i = k+1$ and $j = L$ in (28), then for $0 \leq k \leq L-2$, we obtain

$$D^{\mathrm{p}}_{1,2,\ldots,L-k}(r^*_{k+1}, r^*_{k+2}, \ldots, r^*_L;\ c^*_{k+1}, c^*_{k+2}, \ldots, c^*_L;\ \alpha) < D^{\mathrm{p}}_1(r^*_L;\ c^*_L;\ \alpha), \tag{46}$$

where the range of k is derived from $1 \leq i \leq L-1$ in (28). In addition, from (43), we obtain

$$D^{\mathrm{p}}_1(r^*_L;\ c^*_L;\ \alpha) = \sigma^2 p(r^*_L, c^*_L) + \sigma^2 g\big(b(r^*_L, c^*_L)\big)\big(1 - p(r^*_L, c^*_L)\big) < \sigma^2, \tag{47}$$

where the inequality follows from $p(r^*_L, c^*_L) < 1$ and $0 < g\big(b(r^*_L, c^*_L)\big) < 1$. From (45), (46) and (47), it can be shown that (37) holds for $1 \leq i \leq L-2$ and $j = L$. We next consider the case of $i = L-1$ and $j = L$. From (45) and (47), it is clear that (37) holds for $i = L-1$ and $j = L$. We have thus shown that (37) holds for $1 \leq i \leq L-1$ and $j = L$. Finally, from i) and ii), it is seen that the condition of this theorem implies that (37) is valid for $1 \leq i \leq L-1$ and $i+1 \leq j \leq L$.

In the following, we will prove that if (37) is satisfied, (34) holds for some integers i, j in the range of $1 \leq i \leq L-1$ and $i+1 \leq j \leq L$, by contradicting the following assumption: Eq. (37) holds with at least one of the following three conditions: i) $r^*_i > r^*_j,\ c^*_i > c^*_j,$ ii) $r^*_i > r^*_j,$ $c^*_i = c^*_j,$ iii) $r^*_i = r^*_j,\ c^*_i > c^*_j$. From (1) and (2) together with at least one of the conditions above, it can be readily shown that

$$p(r^*_i, c^*_i) > p(r^*_j, c^*_j). \tag{48}$$

Further, from the fact that $g(x) = 2^{-\alpha x}\ (\alpha \geq 2)$ and $b(u_i, v_i) = u_i\, v_i\, T_{\mathrm{pkt}} W_{\mathrm{pkt}} > 0$, we have

$$g\big(b(r^*_i, c^*_i)\big) < g\big(b(r^*_j, c^*_j)\big). \tag{49}$$

It is clear from (48) and (49) that (37) is not satisfied, and hence the assumption is false. We have thus shown that the condition of this Theorem implies that (34) holds for some integers i, j in the range of $1 \leq i \leq L-1$ and $i+1 \leq j \leq L$. Letting $j = i+1, i+2, \ldots, L$ in (34), we obtain at least $L-i$ constraints on r^*_i or c^*_i of the ith packet.

□

Corollary 6 follows immediately from Theorem 5.

Corollary 6: If $\mathbf{s}_{1,2,\ldots,L} = [r_1^*, r_2^*, \ldots, r_L^*;\ c_1^*, c_2^*, \ldots, c_L^*]$ is a parametric distortion-based optimum for L progressive packets, i.e.,

$$\begin{aligned} D^{\mathrm{p}}_{1,2,\ldots,L}(r_1^*, r_2^*, \ldots, r_L^*;\ c_1^*, c_2^*, \ldots, c_L^*;\ \alpha) &\leq D^{\mathrm{p}}_{1,2,\ldots,L}(r_1, r_2, \ldots, r_L;\ c_1, c_2, \ldots, c_L;\ \alpha) \\ \text{for any } r_1, r_2, \ldots, r_L \in \mathcal{R} &\text{ and } c_1, c_2, \ldots, c_L \in \mathcal{C}, \end{aligned} \tag{50}$$

then, for every integer i, j in the range of $1 \leq i \leq L-1$ and $i+1 \leq j \leq L$, at least one of the following three conditions holds:

$$\text{i) } r_i^* < r_j^*, \quad \text{ii) } c_i^* < c_j^*, \quad \text{iii) } r_i^* = r_j^*,\ c_i^* = c_j^*. \tag{51}$$

That is, we obtain at least $(L^2 - L)/2$ constraints on $r_1^*, r_2^*, \ldots, r_{L-1}^*$ or $c_1^*, c_2^*, \ldots, c_{L-1}^*$.

Proof: The condition of this corollary, given by (50), is identical to (17) of Lemma 2 when we let $i = 1$ in (17). Thus, (18) of Lemma 2 holds for some integer k in the range of $2 \leq k \leq L$ as follows:

$$\begin{aligned} &D^{\mathrm{p}}_{1,2,\ldots,L-k+1}(r_k^*, r_{k+1}^*, \ldots, r_L^*;\ c_k^*, c_{k+1}^*, \ldots, c_L^*;\ \alpha) \\ &\leq D^{\mathrm{p}}_{1,2,\ldots,L-k+1}(r_k, r_{k+1}, \ldots, r_L;\ c_k, c_{k+1}, \ldots, c_L;\ \alpha) \\ &\text{for any } r_k, r_{k+1}, \ldots, r_L \in \mathcal{R} \text{ and } c_k, c_{k+1}, \ldots, c_L \in \mathcal{C}. \end{aligned} \tag{52}$$

If we let $i = k$ in the condition of Theorem 5, given by (17), then it equals (52) in the range of $2 \leq k \leq L-1$. Note that this range of k is a subset of $1 \leq k \leq L-1$ and $2 \leq k \leq L$ given by (17) and (52), respectively. As a result, it follows from Theorem 5 that, for $2 \leq k \leq L-1$ and $k+1 \leq j \leq L$, (51) holds with k being substituted into i. In addition, from Theorem 5 and (50), it follows immediately that for $2 \leq j \leq L$, (51) holds with $i = 1$. We have thus shown that (51) holds for $1 \leq i \leq L-1$ and $i+1 \leq j \leq L$. Letting $i = 1, 2, \ldots, L-1$ and $j = i+1, i+2, \ldots, L$ in (51), we obtain at least $(L^2 - L)/2$ $(= \sum_{i=1}^{L-1} L - i)$ constraints on $r_1^*, r_2^*, \ldots, r_{L-1}^*$ or $c_1^*, c_2^*, \ldots, c_{L-1}^*$ of all the L packets except the last one.

□

By Corollary 6, we are able to reduce the search space of $\mathcal{R}$ and $\mathcal{C}$ when finding a parametric distortion-based optimal solution. Corollary 6 implies that for $1 \leq i < j \leq L$, we possibly have $r_i^* > r_j^*$ if the condition of $c_i^* < c_j^*$, given by (51), is satisfied. That is, it is not guaranteed that r_i^* is nondecreasing in the packet number i $(1 \leq i \leq L)$. This differs from the typical unequal error protection strategy in a SISO system, where a larger spectral efficiency is usually assigned

to the packet later in the sequence, due to the steadily decreasing importance for bits later in the progressive bitstream.

Based on Theorem 1 and Corollary 6, a parametric distortion-based optimal solution for L progressive packets can be obtained in the following way.

Step 1: Choose the best parameter α^* of the parametric model of the distortion-rate function, $f^{\mathrm{p}}(x) = \sigma^2 2^{-\alpha x}$, as follows:

$$\alpha^* = \underset{\alpha \in \{\alpha_1, \alpha_2, \ldots, \alpha_Q\}}{\arg\min} \; D_{1,2,\ldots,L}(r_1^*(\alpha), r_2^*(\alpha), \ldots, r_L^*(\alpha);\; c_1^*(\alpha), c_2^*(\alpha), \ldots, c_L^*(\alpha)), \quad (53)$$

where $D_{1,2,\ldots,L}(\cdot)$ is the expected distortion of the source, given by (4), employing the actual distortion-rate function $f(x)$. For a given parameter $\alpha \in \{\alpha_1, \alpha_2, \ldots, \alpha_Q\}$, $r_i^*(\alpha)$ and $c_i^*(\alpha)$ can be obtained through Steps 2–4 below.

From α^* chosen in (53), we obtain a parametric distortion-based optimal solution of $r_1^*(\alpha^*)$, $r_2^*(\alpha^*)$, …, $r_L^*(\alpha^*)$ and $c_1^*(\alpha^*)$, $c_2^*(\alpha^*)$, …, $c_L^*(\alpha^*)$.

Step 2: In order to find $r_i^*(\alpha)$ and $c_i^*(\alpha)$ for a given parameter α, the packet index i is initialized as $i = L$.

Step 3: If $i = L$,

$$r_L^*(\alpha),\; c_L^*(\alpha) = \underset{r_L \in \mathcal{R},\; c_L \in \mathcal{C}}{\arg\min} \; D_1^{\mathrm{p}}(r_L;\; c_L;\; \alpha), \quad (54)$$

otherwise (i.e., $1 \leq i \leq L-1$),

$$\begin{aligned} & r_i^*(\alpha),\; c_i^*(\alpha) = \\ & \underset{r_i \in \mathcal{R},\; c_i \in \mathcal{C}}{\arg\min} \; D^{\mathrm{p}}_{1,2,\ldots,L-i+1}(r_i, r_{i+1}^*(\alpha), \ldots, r_L^*(\alpha);\; c_i, c_{i+1}^*(\alpha), \ldots, c_L^*(\alpha);\; \alpha), && (55) \\ & \qquad \text{subject to at least one of the three constraints : i) } r_i^*(\alpha) < r_j^*(\alpha); \\ & \qquad \text{ii) } c_i^*(\alpha) < c_j^*(\alpha);\; \text{iii) } r_i^*(\alpha) = r_j^*(\alpha),\; c_i^*(\alpha) = c_j^*(\alpha),\; \text{for every integer } j \text{ in} \\ & \qquad \text{the range of } i+1 \leq j \leq L, && (56) \end{aligned}$$

where $D^{\mathrm{p}}_{1,2,\ldots,L-i+1}(\cdot)$ $(1 \leq i \leq L)$ is the expected distortion of the source with

parametric distortion-rate function $f^{\mathrm{p}}(x)$, and is given by

$$
\begin{aligned}
&D^{\mathrm{p}}_{1,2,\ldots,L-i+1}(r_i, r^*_{i+1}(\alpha), \ldots, r^*_L(\alpha);\ c_i, c^*_{i+1}(\alpha), \ldots, c^*_L(\alpha);\ \alpha) = \\
&\sigma^2 p(r_i, c_i) + \sigma^2 g\big(b(r_i, c_i)\big)\big(1 - p(r_i, c_i)\big)\left[\sum_{n=1}^{L-i}\left(\prod_{k=2}^{n} g\big(b(r^*_{i+k-1}(\alpha), c^*_{i+k-1}(\alpha))\big)\right)\right. \\
&\times p(r^*_{i+n}(\alpha), c^*_{i+n}(\alpha)) \prod_{k=2}^{n}\big(1 - p(r^*_{i+k-1}(\alpha), c^*_{i+k-1}(\alpha))\big) \\
&\left. + \prod_{k=2}^{L+1-i} g\big(b(r^*_{i+k-1}(\alpha), c^*_{i+k-1}(\alpha))\big) \prod_{k=2}^{L+1-i}\big(1 - p(r^*_{i+k-1}(\alpha), c^*_{i+k-1}(\alpha))\big)\right]. \qquad (57)
\end{aligned}
$$

Step 4: Set $i = i - 1$. If $i = 0$, we have obtained $r^*_1(\alpha)$, $r^*_2(\alpha)$, $\ldots$, $r^*_L(\alpha)$, and $c^*_1(\alpha)$, $c^*_2(\alpha)$, $\ldots$, $c^*_L(\alpha)$ for a given parameter α; thus go to Step 1. Otherwise, go to Step 3. $\square$

We first describe Steps 2–4. Eq. (55) follows from Theorem 1. In (54) and (55), it is shown that the joint assignment of spectral efficiencies and spatial multiplexing rates to L packets is optimized in a packet-by-packet manner; that is, for the ith packet, only two optima $r^*_i(\alpha)$ and $c^*_i(\alpha)$ are exhaustively searched ($1 \le i \le L$), from which the global minimum of $D^{\mathrm{p}}_{1,2,\ldots,L}(\cdot)$ can be attained. Eq. (57) in Step 3 can be derived from (13). From (54), (55), and (57), it is seen that the selection of $r^*_i(\alpha)$ and $c^*_i(\alpha)$ does not depend on how large $\sigma^2\,(\neq 0)$ is, but depends on how large $\alpha\,(\ge 2)$ is. To emphasize this, in Steps 1–4, we have used the notation $r^*_i(\alpha)$ and $c^*_i(\alpha)$ instead of r^*_i and c^*_i, respectively. The constraint given by (56), which reduces the search space of $\mathcal{R}$ and $\mathcal{C}$ for the evaluation in (55), is based on Corollary 6. Consequently, following Steps 2–4, the number of ways to assign N_{se} spectral efficiencies and N_{smr} spatial multiplexing rates to L packets, of which the expected distortions need to be evaluated for the optimization, is given by $N_{\mathrm{steps}\,2-4} \le N_{\mathrm{se}} N_{\mathrm{smr}} L$, where the inequality follows from the constraint given by (56). Note that the number of possible assignments for an exhaustive search is $(N_{\mathrm{se}} N_{\mathrm{smr}})^L$.

We next describe Step 1. In (53), the best parameter α^* is chosen to minimize the expected distortion of the source. Note that the actual distortion-rate function $f(x)$ has been used to compute the expected distortion. In (53), due to the high nonlinearity of $f(x)$, we resort to an exhaustive search of α^* that minimizes the expected distortion. If $\alpha \ge 2$ is quantized into Q levels for the exhaustive search, Steps 2–4 should be repeated Q times. This is because for every $\alpha \in \{\alpha_1, \alpha_2, \ldots, \alpha_Q\}$, $r^*_i(\alpha)$ and $c^*_i(\alpha)$ $(1 \le i \le L)$ should be obtained to calculate $D_{1,2,\ldots,L}(\cdot)$ in (53). Thus, the number of ways of assignment, of which the expected distortions need to be

evaluated in Steps 1–4, can be expressed as

$$N_{\text{steps }1-4} \leq QN_{\text{se}}N_{\text{smr}}L, \tag{58}$$

where the inequality follows from the constraint of (56). Eq. (58) indicates that even in the worst case, the number of ways of assignment decreases exponentially from $(N_{\text{se}}N_{\text{smr}})^L$ to $QN_{\text{se}}N_{\text{smr}}L$, relative to an exhaustive search. We note that, however, a parametric distortion-based solution presented in this section is suboptimal in terms of the expected distortion performance, while an exhaustive search obviously yields the optimal performance. In the next section, we will assess the performance of our solution.

IV. Numerical Evaluation

We numerically evaluate the performance of the proposed optimization method presented in Section III. We take three space-time codes into consideration: V-BLAST with an MMSE receiver, OSTBC with a decorrelator, and two-layer D-BLAST with a group zero-forcing receiver [27]. Group decoding is a recent decoding method [28] which can be regarded as a compromise between zero-forcing and maximum likelihood decoding. The spatial multiplexing rates of OSTBC, two-layer D-BLAST, and V-BLAST are denoted by C_1, C_2, and C_3, respectively, and a set of candidate spatial multiplexing rates is given by $\mathcal{C} = \{C_1, C_2, C_3\}$. We have $C_2 = 2N_{\text{t}}/(N_{\text{t}}+1)$ and $C_3 = N_{\text{t}}$ for N_{t} transmit antennas. In the complex OSTBC, the Alamouti code achieves $C_1 = 1$ for $N_{\text{t}} = 2$, and $C_1 = 3/4$ is the maximum achievable rate for $N_{\text{t}} = 3$ or 4. For $N_{\text{t}} \geq 5$, $C_1 = 1/2$ is the maximum rate. We assume a MIMO system with $N_{\text{r}} \geq N_{\text{t}} \geq 2$. The strict inequality of $C_1 < C_2 < C_3$ can be shown to hold for $N_{\text{t}} \geq 2$. In this evaluation, as an example, a set of candidate spectral efficiencies is chosen as $\mathcal{R} = \{1, 2, 3, 4\}$ (bits/s/Hz).

The channel is assumed to experience slow fading such that the fading coefficients remain nearly constant over a packet and are i.i.d. across different packets. We assume that the channel estimation at the receiver is perfect in such slow fading channels. With suitably powerful channel codes, the error probability when not in outage is very small, and hence the outage probability is an accurate approximation of the actual probability of the packet error [29]. For a given spectral efficiency $R \in \mathcal{R}$, the outage probability in a SISO system, denoted by $P_{\text{out, siso}}(\gamma_{\text{s}})$, is calculated from the mutual information of the channels and is given by $P_{\text{out, siso}}(\gamma_{\text{s}}) = \Pr\left[\log_2\left(1 + \gamma_{\text{s}}|h|^2\right) < R\right]$, where h is the complex channel coefficient $\sim \mathcal{CN}(0, 1)$. Let $P_{\text{out},1}(\gamma_{\text{s}})$,

$P_{\mathrm{out},2}(\gamma_{\mathrm{s}})$, and $P_{\mathrm{out},3}(\gamma_{\mathrm{s}})$ denote the outage probabilities of OSTBC, D-BLAST, and V-BLAST, respectively. For a given spectral efficiency $R \in \mathcal{R}$ and a spatial multiplexing rate $C_1 \in \mathcal{C}$, the outage probability of OSTBC is given by [18]

$$P_{\mathrm{out},1}(\gamma_{\mathrm{s}}) = \Pr\left[C_1 \log_2\left(1 + \frac{\gamma_{\mathrm{s}}}{C_1}\|\mathbf{H}\|_F^2\right) < C_1\, R\right] = \Pr\left[\log_2\left(1 + \frac{\gamma_{\mathrm{s}}}{C_1}\|\mathbf{H}\|_F^2\right) < R\right], \quad (59)$$

where $\|\cdot\|_F$ denotes the Frobenius norm of the matrix. The outage probability of D-BLAST, for a given spectral efficiency $R \in \mathcal{R}$ and a spatial multiplexing rate $C_2 \in \mathcal{C}$, is expressed as [27]

$$P_{\mathrm{out},2}(\gamma_{\mathrm{s}}) = \Pr\left[\frac{1}{T_{\mathrm{s}}}\log_2\det(\mathbf{I}_{N_{\mathrm{t}}} + N_{\mathrm{t}}\gamma_{\mathrm{s}}\mathbf{H}_1^H\mathbf{Q}_1^H\mathbf{Q}_1\mathbf{H}_1) < \frac{C_2\, R}{2}\right], \quad (60)$$

where $\mathbf{H}_1$ and $\mathbf{Q}_1$ are given by Eqs. (4) and (6) of [27], respectively. For the V-BLAST scheme, we consider pure spatial multiplexing [30] where data is split into several substreams, one for each transmit antenna, and each substream undergoes independent temporal coding to avoid complex joint decoding of substreams at the receiver. For this scheme, an outage event occurs when any of the substreams is in outage. Thus, the outage probability of V-BLAST, for a given spectral efficiency $R \in \mathcal{R}$ and a spatial multiplexing rate $C_3 \in \mathcal{C}$, is given by [30]

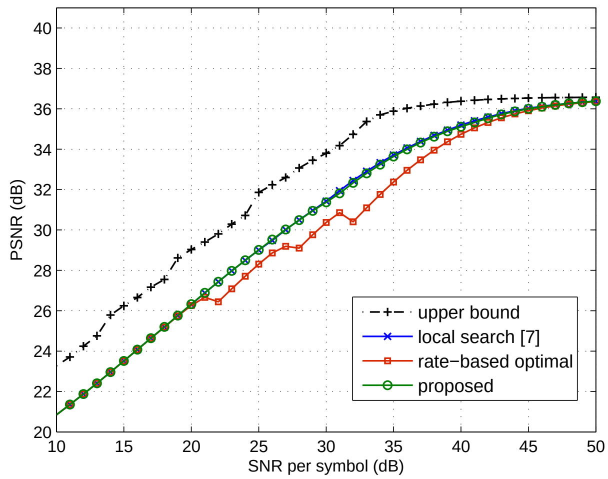
$$P_{\mathrm{out},3}(\gamma_{\mathrm{s}}) = \Pr\left[\bigcup_{k=1}^{N_{\mathrm{t}}}\left\{\log_2\left(1+\gamma_k\right) < \frac{C_3\, R}{N_{\mathrm{t}}}\right\}\right], \quad (61)$$

where $\gamma_k = 1/[(\mathbf{I}_{N_{\mathrm{t}}} + \gamma_{\mathrm{s}}\mathbf{H}^H\mathbf{H})^{-1}]_{kk} - 1$, and $[\cdot]_{kk}$ indicates the kth diagonal entry of a matrix. From the outage probabilities given by (59)–(61), we calculate the probability of the packet error, $p(R, C_i)$, in (57) ($R \in \mathcal{R}$ and $C_i \in \mathcal{C}$).

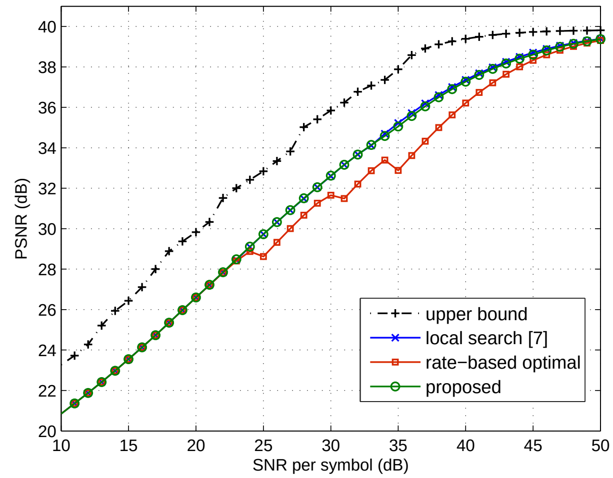
We evaluate the performance of the proposed optimization method using the source coder SPIHT [4] as an example, for an 8 bpp 512×512 Lena image with transmission rates of 0.5 and 1.0 bpp. The number of packets is chosen as $L = 64$ and 128 for the 0.5 and 1.0 bpp rates, respectively.[2] The end-to-end performance is measured by the expected distortion, $D_{1,2,\ldots,L}(\cdot)$. To compare the image quality, we use the peak-signal-to-noise ratio (PSNR) defined as $10\log_{10}(255^2/D_{1,2,\ldots,L}(\cdot))$ (dB). To find the best α^* of the parametric distortion-rate function, $f^{\mathrm{p}}(x) = \sigma^2 2^{-\alpha x}$, we quantize α into Q bins for the range of $2 \leq \alpha \leq M$, such that the width of each bin is given by $(M-2)/Q$. In our evaluation, we set M and Q to be 20 and 1, respectively.

To begin, we observe the PSNR performance of the proposed optimization method when employed in a SISO system. For this case, only spectral efficiencies are optimally assigned to

[2] The transmission rates and the number of packets in this evaluation are the same as those used in [7], [9].



(a) Lena of 0.5 bpp and $L = 64$ packets

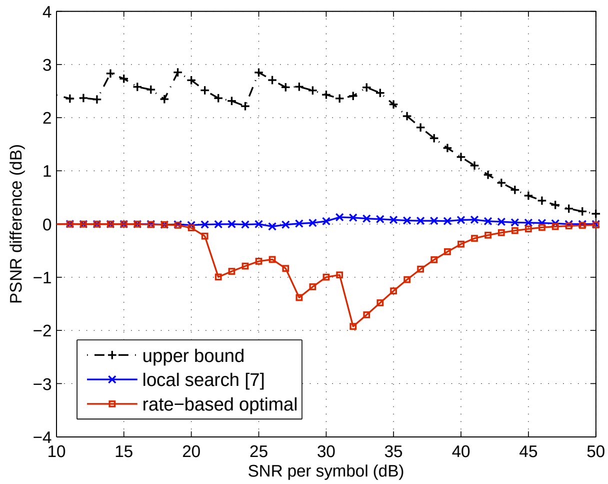


(b) Lena of 1.0 bpp and $L = 128$ packets

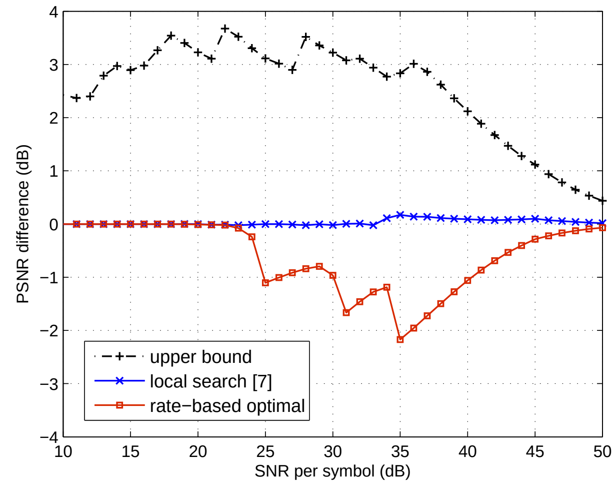
Fig. 2. The PSNR performance for the transmission of 8 bpp 512×512 Lena image in SISO Rayleigh fading channels.

progressive packets, and space-time codes are not employed. Recall that in a SISO system, there have been many studies about the optimal assignment of spectral efficiencies to progressive packets. The local search algorithm [7] and heuristic algorithm based on graph search [9] are among the best optimization methods in terms of the expected distortion performance (or, equivalently, mean squared-error performance). In [9], it is shown that the two algorithms provide nearly identical mean squared-error performances for progressive transmission, which are near optimal [7]. Fig. 2 shows the PSNR performances of the proposed method and the local search algorithm in a SISO system, in addition to showing the upper bound on the PSNR performance (or, equivalently, the lower bound on the expected distortion) that is presented in [7], [9] for reference. Fig. 2 also depicts the PSNR of a rate-based optimal solution [31]–[33] that assigns the spectral efficiencies to packets in a way that the expected number of correctly decoded information bits is maximized, rather than minimizing the expected distortion. The motivation to show the performance of a rate-based optimal solution will be described later. The search space of $\mathcal{R}$ for a sequence of progressive packets is too large to find a global optimal solution through an exhaustive search. Thus, the PSNR achieved by an exhaustive search is not presented in Fig. 2.

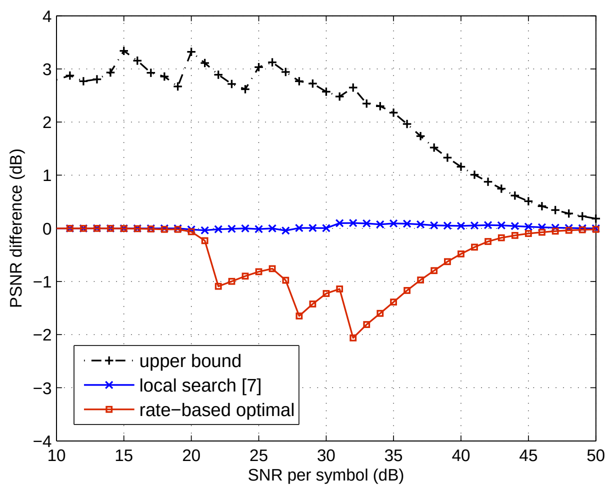
For better visual comparison, in Fig. 3(a) and (b), we depict the PSNR difference between our



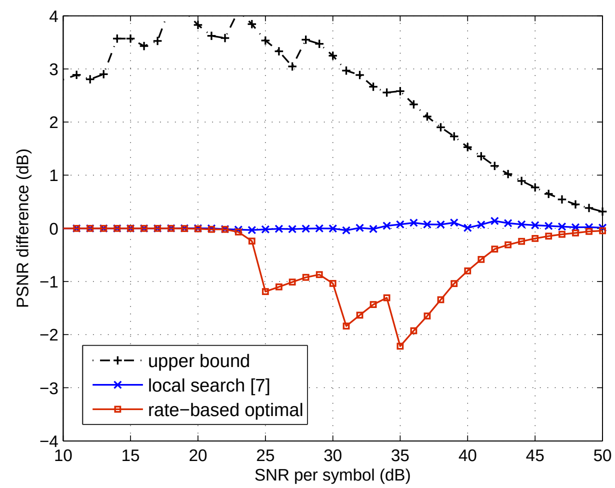
(a) Lena of 0.5 bpp and $L = 64$ packets



(b) Lena of 1.0 bpp and $L = 128$ packets



(c) Pepper of 0.5 bpp and $L = 64$ packets



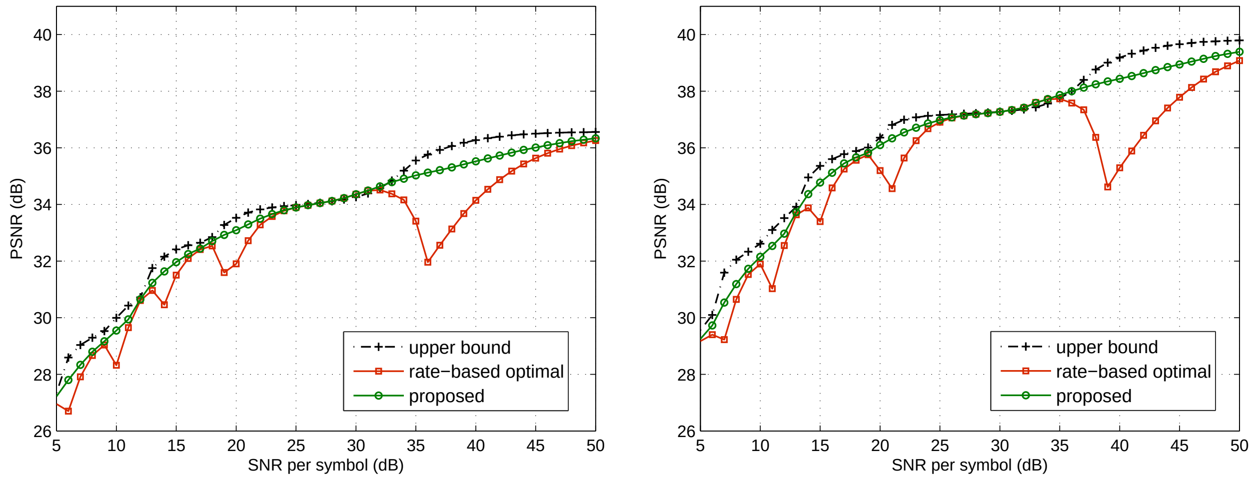
(d) Pepper of 1.0 bpp and $L = 128$ packets

Fig. 3. The PSNRs of various optimization schemes, each subtracted by the PSNR of the proposed method, for the transmission of 8 bpp 512×512 Lena and Pepper images in SISO Rayleigh fading channels.

solution and the others; that is, we show the PSNRs of the upper bound, local search algorithm, and a rate-based optimal solution, each subtracted by the PSNR of our solution. We also observe the PSNR performance for another image, 8 bpp 512×512 Pepper, with rates of 0.5 and 1.0 bpp, and the results are given in Fig. 3(c) and (d). For all of cases in Fig. 3, the PSNR performance of the proposed method is seen to be close to that of the local search algorithm, and is significantly better than that of the rate-based optimal solution.

Next, we evaluate the PSNR performance of the proposed optimization method in a MIMO system. We consider three different MIMO channels: 2×2 i.i.d. Rayleigh fading channels, 3×3 spatially correlated Rayleigh fading channels with correlation coefficients of $\rho_{\mathrm{t}} = \rho_{\mathrm{r}} = 0.7$, and 4×4 Rician fading channels with Rician factor of $K = 5$. The system parameters are set equal to those used in a SISO system. Fig. 4 shows the PSNR of the proposed method in 2×2 i.i.d. Rayleigh fading channels, in addition to showing the PSNR of a rate-based optimal solution and that of the upper bound. Note that in a SISO system, a rate-based optimal assignment of spectral efficiencies to progressive packets, which maximizes the expected number of correctly decoded information bits, can be found in a packet-by-packet manner [33]. This can be immediately extended to a MIMO system such that a rate-based optimal assignment of spectral efficiencies and spatial multiplexing rates to the packets can be obtained using a packet-by-packet method. In other words, we are able to obtain a rate-based optimal solution even for the case where space-time codes are involved in the optimization.

On the other hand, the optimization algorithms that are proposed to minimize the expected distortion in a SISO system do not immediately tell us how to jointly assign the spectral efficiencies and spatial multiplexing rates to progressive packets. For example, in [7]–[9], it is assumed that for a fixed alphabet size of modulation, and for a set of channel code rates given by $k_1 < k_2 < \cdots < k_m$, the corresponding probabilities of the packet error satisfy the inequalities of $p(k_1) < p(k_2) < \cdots < p(k_m)$. That is, a lower probability of the error is implied by a lower data rate, which is defined as the number of information bits in a packet divided by its time duration. This is qualitatively depicted in Fig. 5(a). However, this does not hold in general if the packets are allowed to be encoded using different space-time codes. As an example, in a 2×2 MIMO system, if the spectral efficiency of the packet, encoded by a channel code and a modulation, is R (bits/s/Hz) and if the packet is encoded by V-BLAST, with a spatial multiplexing rate of 2, then the data rate of the packet equals $2R\,W_{\mathrm{pkt}}$ (bits/s). In another case, if the spectral efficiency of the packet is $2R$ and if the Alamouti code, with a spatial multiplexing rate of 1, encodes the packet, then the data rate also equals $2R\,W_{\mathrm{pkt}}$. The work in [15] analytically shows that a crossover point exists for the error probability curves of the above two cases. Further, if the spectral efficiency of the packet encoded with V-BLAST is set lower than R, then the data rate becomes lower than $2R\,W_{\mathrm{pkt}}$, and the resultant error probability curve can be shown to also have a crossover point with that of the packet encoded with the Alamouti code and a



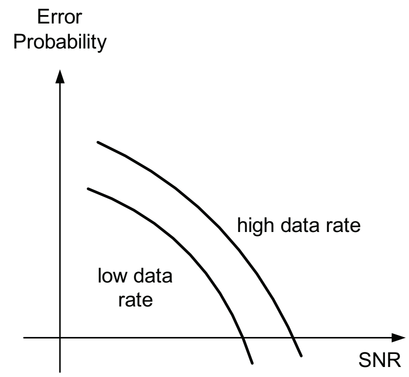
(a) Lena of 0.5 bpp and $L = 64$ packets

(b) Lena of 1.0 bpp and $L = 128$ packets

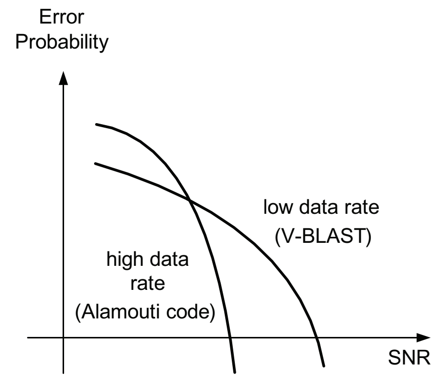
Fig. 4. The PSNR performance for the transmission of 8 bpp 512×512 Lena image in 2×2 i.i.d. Rayleigh fading channels.

spectral efficiency of $2R$. This is because V-BLAST and the Alamouti code offer unequal spatial diversity orders that correspond to unequal asymptotic slopes of the error probability curves [29]. As a result, a packet with a lower data rate does not necessarily yield a lower error probability if packets are encoded using different space-time codes. This is qualitatively depicted in Fig. 5(b). The error probability is not simply an increasing function in the data rate, but rather a more complex function of the parameters including the spatial multiplexing rate and diversity order, which depend on the space-time code and its receiver, as well as the propagation channel characteristics including the spatial correlation and Rician factor.

We are unaware of any algorithm in [7]–[14] that has been successfully extended to the case where space-time coding is also involved in the optimization in a MIMO system. For this reason, we cannot compute and present the PSNR performances of the optimization algorithms that were proposed for a SISO system. Fig. 6 provides a better visual comparison by showing the difference in the PSNR between our solution and the others. For all cases in Fig. 6, the proposed method is able to significantly improve the performance of a rate-based optimal solution. Fig. 6 also shows the performance of a suboptimal case where a space-time code is excluded from a candidate set. It is seen that there is a significant PSNR gap between the two cases where a single space-time code has been excluded and not excluded. This indicates that when progressive



(a) SISO systems



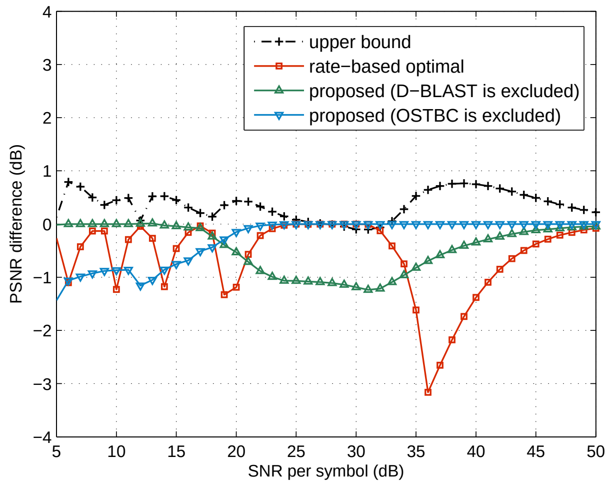
(b) MIMO systems

Fig. 5. (a) SISO systems: a lower data rate implies a lower error probability. (b) MIMO systems: a lower data rate does not necessarily yield a lower error probability.

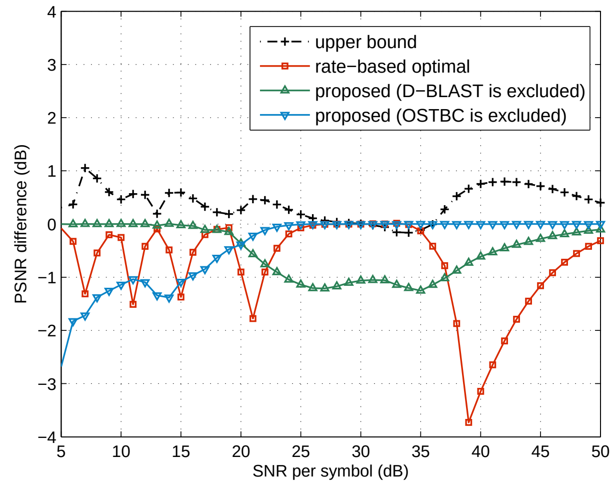
packets are transmitted in a MIMO system, the PSNR performance would improve if a variety of space-time codes are employed to encode a sequence of packets. This motivated us to suggest an optimization method that is able to handle two or more space-time codes to transmit progressive packets in a MIMO system.

For the PSNR performances of the proposed method shown in Figs. 4 and 6, (55) has been computed with the constraint given by (56). We note that the same set of spectral efficiencies and spatial multiplexing rates is obtained when (55) is computed with and without the constraint. That is, the constraint in (56) reduces the computational complexity involved with the optimization without losing any PSNR performance. From Fig. 6, it is observed that the PSNR of the upper bound subtracted by that of our solution is smaller than 0 dB at some points. We conjecture that this is because the derivation of the bound in [7] is based on the hypothesis that the distortion-rate function, $f(x)$, is nonincreasing and convex, which only offers an approximation for practical source coders and images.

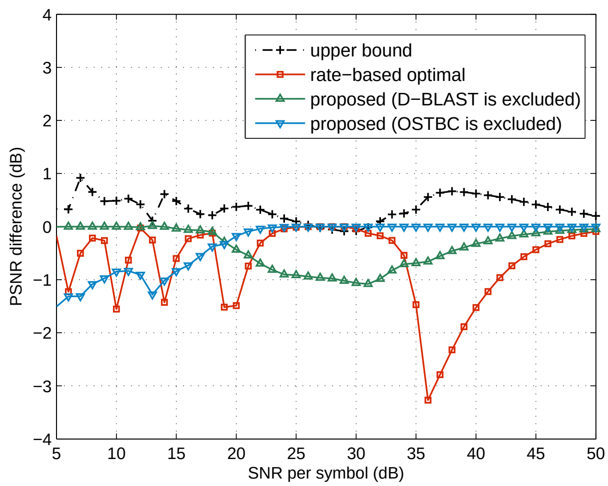
Next, we consider spatially correlated Rayleigh fading and Rician fading channels, instead of i.i.d. Rayleigh fading channels. Note that our optimization method, which is based on Theorem 1 and Corollary 6, is valid over any propagation channels, since no assumption regarding the channel characteristics has been used in the proofs of Theorem 1 and Corollary 6. The numerical results for 3×3 spatially correlated Rayleigh fading channels with $\rho_{\mathrm{t}} = \rho_{\mathrm{r}} = 0.7$, and those for 4×4 Rician fading channels with $K = 5$ are not depicted here due to limited space. However,



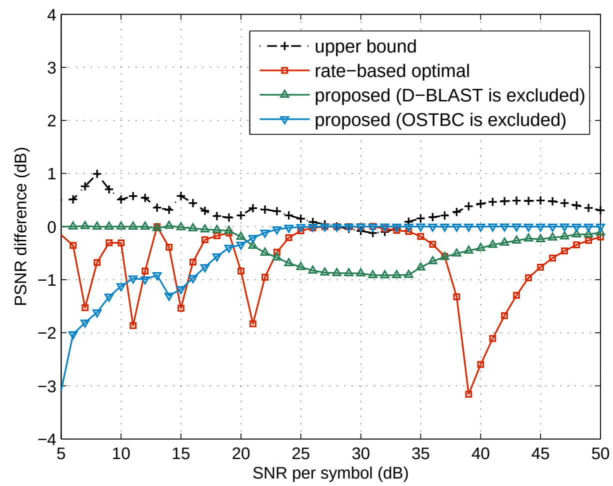
(a) Lena of 0.5 bpp and $L = 64$ packets



(b) Lena of 1.0 bpp and $L = 128$ packets



(c) Pepper of 0.5 bpp and $L = 64$ packets



(d) Pepper of 1.0 bpp and $L = 128$ packets

Fig. 6. The PSNRs of various optimization schemes, each subtracted by the PSNR of the proposed method, for the transmission of 8 bpp 512×512 Lena and Pepper images in 2×2 i.i.d. Rayleigh fading channels.

we note that what we have discussed regarding the numerical results for i.i.d. Rayleigh fading channels is also observed in the results for spatially correlated Rayleigh and Rician fading channels.

V. CONCLUSIONS

The joint optimization of source, channel, and space-time coding for a series of numerous progressive packets is a challenging problem. MIMO systems are incorporated in most modern wireless systems such as 3GPP Long Term Evolution and IEEE 802.16m (WiMAX). Nevertheless, to our knowledge, a feasible solution for the joint optimization problem in a MIMO system has not yet been presented in the literature. This paper uses a parametric methodology to solve such a complex joint optimization problem. In the proposed method, employing a parametric distortion-rate function, we jointly optimize the assignment of spectral efficiencies and spatial multiplexing rates to progressive packets in a packet-by-packet manner. As a result, the computational complexity of the optimization is exponentially reduced, compared to an exhaustive search. Moreover, some constraints on the search space are derived to further reduce the complexity. The numerical results show that the proposed solution significantly improves the PSNR performance of a rate-based optimal solution in a MIMO system. In addition, the performance of our solution when applied to a SISO system is close to that of local search algorithm, one of the best optimization methods proposed for a SISO system.

Lastly, we note that our solution can be computed independently of a specific progressive source, once the best α^* of the parametric function for that source has been chosen. Thus, if a single parameter α^* is known to the receiver side, which requires only a small amount of overhead, our solution for spectral efficiencies and spatial multiplexing rates can then be recomputed at the receiver side. Therefore, the overhead information for the solution is unnecessary. In general, the number of parameters in a parametric model should remain small, because both the overhead and the optimization time grow with the number of parameters. In our approach, only a single parameter α is taken to avoid such a problem.

The work in this paper has significance in terms of its impact on the area of multimedia communications, and deepens our understanding of joint source and channel coding problem in a MIMO system. As a future work, we consider extending our joint source, channel, and space-time coding to progressive transmission in multiview settings where multiple cameras are coded and sent to clients in a progressive fashion.

REFERENCES

[1] H. Zhang, Y. Zheng, M. A. Khojastepour, and S. Rangarajan, “Cross-layer optimization for streaming scalable video over fading wireless networks,” *IEEE J. Select. Areas Commun.*, vol. 28, pp. 344–353, Apr. 2010.

[2] S.-H. Chang, P. C. Cosman and L. B. Milstein, “Iterative channel decoding of FEC-based multiple description codes,” *IEEE Trans. Image Process.*, vol. 21, pp. 1138–1152, Mar. 2012.

[3] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of H.264/AVC,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, pp. 1103–1120, Sept. 2007.

[4] Y. Jin and H.-J. Lee, “A block-based pass-parallel SPIHT algorithm,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, pp. 1064–1075, Jul. 2012.

[5] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, “Space-time block codes from orthogonal designs,” *IEEE Trans. Inf. Theory*, vol. 45, pp. 1456–1467, July 1999.

[6] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multiple antennas,” *Bell Labs. Tech. J.*, vol. 1, pp. 41–59, 1996.

[7] R. Hamzaoui, V. Stankovic, and Z. Xiong, “Fast algorithm for distortion-based error protection of embedded image codes,” *IEEE Trans. Image Process.*, vol. 14, pp. 1417–1421, Oct. 2005.

[8] L. Cao, “On the unequal error protection for progressive image transmission,” *IEEE Trans. Image Process.*, vol. 16, pp. 2384–2388, Sep. 2007.

[9] M. Fresia and F. Lavagetto, “Determination of optimal distortion-based protection in progressive image transmission: A heuristic approach,” *IEEE Trans. Image Process.*, vol. 17, pp. 1654–1662, Sep. 2008.

[10] B. A. Banister, B. Belzer, and T. R. Fischer, “Robust image transmission using JPEG2000 and turbo codes,” *IEEE Signal Process. Lett.*, vol. 9, pp. 117–119, Apr. 2002.

[11] A. Nosratinia, J. Liu, and B. Aazhang, “Source-channel rate allocation for progressive transmission of images,” *IEEE Trans. Commun.*, vol. 51, pp. 186–196, Feb. 2003.

[12] Z. Wu, A. Bilgin, and M. W. Marcellin, “Joint source/channel coding for image transmission with JPEG2000 over memoryless channels,” *IEEE Trans. Image Process.*, vol. 14, pp. 1020–1032, Aug. 2005.

[13] X. Pan, A. H. Banihashemi, and A. Cuhadar, ”A fast trellis-based rateallocation algorithm for robust transmission of progressively coded images over noisy channels,” *IEEE Trans. Commun.*, vol. 54, pp. 1–6, Jan. 2006.

[14] G. Baruffa, P. Micanti, and F. Frescura, “Error protection and interleaving for wireless transmission of JPEG 2000 images and video,” *IEEE Trans. Image Process.*, vol. 18, pp. 346–356, Feb. 2009.

[15] S.-H. Chang, M. Rim, P. C. Cosman and L. B. Milstein, “Superposition MIMO coding for the broadcast of layered sources,” *IEEE Trans. Commun.*, vol. 59, pp. 3240–3248, Dec. 2011.

[16] S. Lin, A. Stefanov, and Y. Wang, “On the performance of space-time block-coded MIMO video communications,” *IEEE Trans. Veh. Technol.*, vol. 56, pp. 1223–1229, May 2007.

[17] T. Holliday, A. J. Goldsmith, and H. V. Poor, “Joint source and channel coding for MIMO systems: Is it better to be robust or quick?,” *IEEE Trans. Inf. Theory*, vol. 54, pp. 1393–1405, Apr. 2008.

[18] S.-H. Chang, P. C. Cosman and L. B. Milstein, “Optimal transmission of progressive sources based on the error probability analysis of SM and OSTBC,” *IEEE Trans. Veh. Technol.*, vol. 63, pp. 94–106, Jan. 2014.

[19] S.-H. Chang, J. P. Choi, P. C. Cosman and L. B. Milstein, “Optimization of multimedia progressive transmission over MIMO channels,” *IEEE Trans. Veh. Technol.*, vol. 65, pp. 1244–1260, Mar. 2016.

[20] S.-H. Chang, J. P. Choi, P. C. Cosman and L. B. Milstein, “Optimization of scalable broadcast for a large number of antennas,” to appear in *IEEE Trans. Veh. Technol.*.

[21] J. Song and K. J. R. Liu, “Robust progressive image transmission over OFDM systems using space-time block code,” *IEEE Trans. Multimedia*, vol. 4, pp. 394–406, Sep. 2002.

[22] D. Song and C. W. Chen, “Scalable H.264/SVC video transmission over MIMO wireless systems with adaptive channel selection based on partial channel information,” in *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, pp. 1218–1226, Sept. 2007.

[23] H. Xiao, Q. Dai, X. Ji, and W. Zhu, “A novel jscc framework with diversity-multiplexing-coding gain tradeoff for scalable video transmission over cooperative mimo,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, pp. 994–1006, July 2010.

[24] S. Zhao, Z. Xiong, X. Wang, and J. Hua, “Progressive video delivery over wideband wireless channels using space-time differentially coded OFDM systems,” *IEEE Trans. Mobile Comput.*, vol. 5, pp. 303–316, Apr. 2006.

[25] G. Taricco and E. Riegler, “On the ergodic capacity of correlated Rician fading MIMO channels with interference,” *IEEE Trans. Inf. Theory*, vol. 57, pp. 4123–4137, Jul. 2011.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

[27] Y. Lu, W. Zhang, X.-G. Xia, “On diversity and multiplexing tradeoff of two-layer D-BLAST with group zero-forcing detection,” *IEEE Trans. Commun.*, vol. 60, pp. 2255–2264, Aug. 2012.

[28] W. Zhang, T. Xu, and Z.-G. Xia, “Two designs of space-time block codes achieving full diversity with partial interference cancellation group decoding,” *IEEE Trans. Inf. Theory*, vol. 58, pp. 747–764, Feb. 2012.

[29] A. Lozano and N. Jindal, “Transmit diversity vs. spatial multiplexing in modern MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 9, pp. 186–197, Jan. 2010.

[30] K. Raj Kumar, G. Caire, and A. L. Moustakas, “Asymptotic performance of linear receivers in MIMO fading channels,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 4398–4418, Oct. 2009.

[31] V. Chande and N. Farvardin, “Progressive transmission of image over memoryless noisy channels,” *IEEE J. Sel. Areas Commun.*, vol. 18, pp. 850–860, Jun. 2000.

[32] A. Hedayat and A. Nosratinia, “Rate allocation in source-channel coding of images,” in *Proc. IEEE Int. Conf. Image Processing*, vol. I, Thessaloniki, Greece, Oct. 2001, pp. 189–192.

[33] V. Stankovic, R. Hamzaoui, and D. Saupe, “Fast algorithm for rate-based optimal error protection of embedded codes,” *IEEE Trans. Commun.*, vol. 51, pp. 1788–1795, Nov. 2003.